THE ANNALS *of* MATHEMATICAL STATISTICS

FOUNDED AND EDITED BY H. C. CARVER, [illegible]
EDITED BY S. S. WILKS, [illegible]
THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

(Continued on back cover)

Vol. 31, No. 3 — September, 1960

THE ANNALS OF MATHEMATICAL STATISTICS

Subscription Rates. Current issues are $15 per volume (four issues of one calendar year) in the U. S. and Canada, $10 per volume elsewhere. Single issues are $4. Back numbers for all issues up to and including 1956 (Vol. XXVII) are $12 per volume, $3.50 per issue, $200 for the first 25 volumes, $10 per additional volume purchased at the same time as Volume I through XXV.

Rates to members of the Institute of Mathematical Statistics are lower (see inside back cover).

Communications concerning subscriptions, back numbers, payment of dues, etc., should be addressed to Gerald J. Lieberman, Treasurer, Institute of Mathematical Statistics, Department of Statistics, Stanford University, Stanford, Calif.

Communications concerning membership, changes of address, etc., should be addressed to George E. Nicholson, Jr., Secretary, Department of Statistics, University of North Carolina, Chapel Hill, N. C. Changes of address which are to become effective for a given issue of the *Annals* should be reported to the Secretary on or before the 10th of the month preceding the month of issue.

Editorial Office, Department of Statistics, Eckhart Hall, University of Chicago, Chicago 37, Illinois. William Kruskal, Editor.

Preparation of manuscripts. Manuscripts should be submitted to the editorial office. Each manuscript should be typewritten, double spaced, with wide margins at sides, top, and bottom, and the original should be submitted with one additional copy, on paper that will take corrections. Dittoed or mimeographed papers are acceptable only if completely legible. Footnotes should be reduced to a minimum, and where possible replaced by remarks in the text, or a bibliography at the end of the paper; formulae in footnotes should be avoided. References should follow current *Annals* style, and should be numbered alphabetically according to authors' names.

Figures, charts, and diagrams should be professionally drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed.

Authors are asked to keep in mind the typographical difficulties of complicated mathematical formulae. The difference between capital and lower-case letters should be clearly shown; care should be taken to avoid confusion between such pairs as zero and the letter O, the numeral 1 and the letter l, numeral 1 used as superscript and prime ('), alpha and a, kappa and k, mu and u, nu and v, eta and n, etc. Subscripts or superscripts should be clearly below or above the line. Bars above groups of letters (e.g., $\overline{\log x}$) and underlined letters (e.g., $\underline{x}$) are difficult to print and should be avoided. Symbols are automatically italicized by the printer and should not be underlined on manuscripts. Boldface letters may be indicated by underlining with a wavy line on the manuscript; boldface subscripts and superscripts are not available. Complicated exponentials should be represented with the symbol exp. In writing square roots the fractional exponent is preferable to the radical sign. Fractions *in the body of the text* (as opposed to displayed expressions) and fractions occurring in the numerators or denominators of fractions are preferably written with the solidus; thus $(a + b)/(c + d)$ rather than $\frac{a+b}{c+d}$.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers will be furnished at cost.

Mail to the *Annals of Mathematical Statistics* should be addressed to either the Editor or the Treasurer, as described above. It should not be addressed to Waverly Press.

COMPOSED AND PRINTED AT THE

WAVERLY PRESS, INC., BALTIMORE, MARYLAND, U. S. A.

Second-class postage paid at Baltimore, Maryland

Published quarterly by the Institute of Mathematical Statistics in March, June, September and December.

(Continued on next page)

Southern Methodist University, Mathematics Department, Dallas 5, Texas
Space Technology Laboratories, P. O. Box 95001, Los Angeles 45, California
Stanford University, Girshick Memorial Library, Stanford, California
State University of Iowa, Iowa City, Iowa
The Atlantic Refining Company, 2700 Passyunk Avenue, Philadelphia, Pa.
Union Carbide Corporation, 30 East 42nd Street, New York 17, New York
Union Oil Company of California, Union Research Center, Box 76, Brea, California
United States Steel Corporation Library, Monroeville, Penna.
University of California, Statistical Laboratory, Berkeley, California
University of Illinois, Serials Department, Urbana, Illinois
University of North Carolina, Department of Statistics, Chapel Hill, North Carolina
University of Puerto Rico, School of Tropical Medicine, San Juan, Puerto Rico
University of Washington, Laboratory of Statistical Research, Seattle, Washington
W. R. Grace and Company, Research Division, Washington Research Center, Clarksville, Maryland
W. R. Grace and Company, Dewey and Almy Chemical Division, 62 Whittemore Avenue, Cambridge 40, Massachusetts

Werner Gautschi, 1927–1959

By J. R. Blum

Sandia Corporation

Werner Gautschi was born on December 11, 1927, in Basel. A serious heart ailment suffered as a young boy prevented him from participating in many of the usual childhood activities and led to an early devotion to mathematics and music. In 1946 he entered the University of Basel and remained there until 1952, with the exception of three terms at Cambridge University during 1950–51. He graduated *summa cum laude* from the University of Basel in 1952, with a dissertation written under the direction of Professor A. Ostrowski.

An early interest in Statistics and Computing brought him to the United States in 1953 in order to study these fields. He spent his first year here at the Institute for Advanced Studies, where he did computational work on eigenvalues and norms of matrices. In 1954 he joined the Statistical Laboratory at Berkeley for a two year period. Aside from his studies, research, and teaching, he made many valuable suggestions to Erich Lehmann who was writing *Testing Statistical Hypotheses* and to Henry Scheffé who was writing *The Analysis of Variance.*

In the fall of 1956 he joined the faculty of Ohio State University and in the fall of 1957 he came to Indiana University for a two year period. During the summer of 1958 he returned to Switzerland where he married Erika Wüst and brought her back to the United States. In the summer of 1959 he rejoined Ohio State University where he remained until his death on October 3, 1959. A son, Thomas, was born on January 25, 1960.

The death of a good man is a loss to all of us. Werner Gautschi was a good man, a fine scientist, and a sensitive pianist. His many friends and colleagues mourn him and remember him.

Bibliography of Werner Gautschi

[1] "The asymptotic behaviour of powers of matrices," *Duke Math. J.* Vol. 20 (1953), pp. 127–140.

[2] "The asymptotic behaviour of powers of matrics II," *Duke Math. J.* Vol. 20 (1953), pp. 375–379.

[3] "Bounds of matrices with regard to an Hermitian metric," *Compositio Math.* Vol. 12 (1954), pp. 1–16.

[4] "Some remarks on systematic sampling," *Ann. Math. Stat.* Vol. 28 (1957), pp. 385–394.

[5] "Some remarks on Herbach's paper, 'Optimum nature of the *F*-test for model II in the balanced case,'" *Ann. Math. Stat.*, Vol. 30 (1959), pp. 960–963.

Received April 7, 1960.

THE CAPACITIES OF CERTAIN CHANNEL CLASSES UNDER RANDOM CODING[1]

By David Blackwell, Leo Breiman, and A. J. Thomasian

University of California, Berkeley

1. Introduction and Summary. For any two finite sets U, V, a Markov matrix s with row set U and column set V will be called a U, V *channel*. Thus a U, V channel is any nonnegative function s, defined for all pairs (u, v), $u \varepsilon U$, $v \varepsilon V$, for which

$$\sum_v s(u, v) = 1 \qquad \text{for all } u.$$

The sets U, V will be called the *input* and *output* sets, respectively, of the channel. We shall denote by $M(U, V)$ the set of all U, V channels. A channel s may be thought of as a random device which, on being given an input element $u \varepsilon U$, produces an output element $v \varepsilon V$, with the probability of a particular output v given by $s(u, v)$.

A U, V channel s may be used as a means of communication from one person, the sender, to another person, the receiver. There is given in advance a finite set D of messages, exactly one of which will be presented to the sender for transmission. The sender encodes the message by an *encoding channel* $s_1 \varepsilon M(D, U)$, with $s_1(d, u)$ being the probability that input u is given to channel s when message d is presented to the sender for transmission. When the receiver observes the output v of the transmission channel s, he decodes it by a *decoding channel* $s_2 \varepsilon M(V, D)$, with $s_2(v, d)$ being the probability that, on receiving the transmission channel output v, the receiver will decide that message d is intended. The pair (s_1, s_2) will be called a (D, U, V) code. For a U, V channel s and a (D, U, V) code $c = (s_1, s_2)$, the matrix $\epsilon(s, c) = s_1 s s_2$, which is an element of $M(D, D)$ will be called the *error matrix* of code c in channel s. Its (d, d') element is the probability that, when message d is presented to the sender, the receiver will decide that message d' is intended, when code c is used on channel s. We shall be especially interested in the *average error probability* over all messages in the set D. This is the number

$$\pi(s, c) = 1 - |D|^{-1} \text{ trace } \epsilon(s, c),$$

where $|D|$ denotes the number of elements in the set D.

A code $c = (s_1, s_2)$ will be called *pure* if only 0's and 1's occur as elements of s_1, s_2. The (finite) set of all pure (D, U, V) codes will be denoted by $C(D, U, V)$, and a probability distribution k over $C(D, U, V)$ will be called a *random*

Received October 21, 1959.

[1] This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-53). This paper in whole or in part may be reproduced for any purpose of the United States Government.

(D, U, V) *code*. We define the error matrix $\epsilon(s, k)$ and average error probability $\pi(s, k)$ for a random code k by

$$\epsilon(s, k) = \sum_{c \varepsilon C(D,U,V)} k(c)\, \epsilon(s, c), \qquad \pi(s, k) = 1 - |D|^{-1} \text{ trace } \epsilon(s, k).$$

It was observed by Shannon [4] that every (D, U, V) code c is equivalent to some random (D, U, V) code k, in the sense that

$$\epsilon(s, k) = \epsilon(s, c) \qquad \text{for all } s \varepsilon M(U, V).$$

The converse is not true. The greater generality of random codes lies in the possibility, with random codes, of correlated randomization in the encoding and decoding processes. This is a special case of the fact in game theory, noted by Kuhn [3], that every behavior strategy (code) is equivalent to some mixed strategy (random code), but the converse holds only in games of perfect recall (which the communication game is not).

Shannon's basic work in information theory [5], and most later work, has been concerned with the question: for a given U, V channel s and message set D, is there a pure code c which makes the average error probability $\pi(s, c)$ (or the maximum error probability) small? For this question, the distinction between pure codes and random codes is irrelevant (though even here random codes are useful as tools [5]), since

$$\pi(s, k) = \sum k(c)\pi(s, c),$$

so that there is a pure code whose average error probability is at least as small as that of any random code. We shall be concerned with some cases in which D, U, V are given, but the transmission channel is known only to be some U, V channel in a given closed set $S \subset M(U, V)$. We ask: is there a random code k for which $\pi(s, k)$ is small for every $s \varepsilon S$? For this question, as we shall see, the distinction between random codes and pure codes is essential, for some sets S.

Specifically, we shall be interested in D, U, V, S defined as follows. We are given a message set D (only $|D|$, the number of elements in D, will be relevant), an input set A, an output set B, a closed set S_0 of A, B channels, and a positive integer N. The sender will be given some message d from D, and will then choose a sequence $u = (a_1, \cdots, a_N)$ of N elements of A. These inputs will be placed successively into channels $s_1, \cdots, s_N$, $s_n \varepsilon S_0$, and the receiver will observe the resulting output sequence $v = (b_1, \cdots, b_N)$. The receiver must then estimate which message d was presented to the sender. Thus U is the set of all sequences $u = (a_1, \cdots, a_N)$ of length N of elements of A and V is the set of all sequences $v = (b_1, \cdots, b_N)$ of length N of elements of B. The set S of possible U, V channels will depend on what restrictions we place on the sequences $s_1, \cdots, s_N$. We consider three cases.

CASE 1. *Fixed unknown channel.* Here we are given that the same element of S_0 is the transmission channel for each period. There is then one U, V channel s for each A, B channel $s_0 \varepsilon S_0$. The s corresponding to s_0 is defined by

$$s(u, v) = \prod_{n=1}^{N} s_0(a_n, b_n).$$

We shall denote the set of all such U, V channels by S_1.

CASE 2. *Arbitrarily varying channel.* Here there is one U, V channel for each sequence $(s_1, \cdots, s_N)$ of elements of S_0, defined by

$$s(u, v) = \prod_{n=1}^{N} s_n(a_n, b_n).$$

We shall denote the set of all such U, V channels by S_2.

CASE 3. *Channel selected by jammer with knowledge of past inputs and outputs.* Here we suppose that the element s_n of S_0 which will be the transmission channel during the nth period if selected by a jammer after he has observed the inputs $a_1, \cdots, a_{n-1}$ and outputs $b_1, \cdots, b_{n-1}$ during the first $n - 1$ periods. A pure strategy f for the jammer is a sequence $(f_1, \cdots, f_N)$ of functions, where f_n maps every sequence $x_n = (a_1, \cdots, a_{n-1}, b_1, \cdots, b_{n-1})$ into a corresponding element $f_n(x_n)$ of S_0. There is then one U, V channel s_f for each pure strategy f, defined by

$$s_f(u, v) = \prod_{n=1}^{N} s_n(a_n, b_n), \qquad \text{where } s_n = f_n(x_n).$$

We shall denote the set of all such U, V channels by S_3.

Let us define, for $i = 1, 2, 3$,

$$\pi_i(|D|, N, S_0) = \min_k \max_{s \in S_i} \pi(s, k).$$

The number $\pi_i(|D|, N, S_0)$ is the minimum average error probability which can be guaranteed, by using a suitable random code, when there are $|D|$ possible messages, N transmission periods, the channel at each period is some element of S_0, and the channel variation from period to period is as described in Case i above. It is also the value of the following two-person zero sum game: Player I (the jammer) chooses any U, V channel s in S_i, and Player II independently chooses a pure (D, U, V) code c. A message is then selected at random from D, so that each d has probability $|D|^{-1}$ of being selected, and transmitted over channel s using code c. If an error is made, Player I wins one unit; otherwise he wins zero.

Since π is linear in s, we have

$$\pi_i(|D|, N, S_0) = \min_k \max_{s \in S_i^*} \pi(s, k),$$

where S_i^* is the convex hull of S_i, i.e. the smallest convex set containing S_i. Let us for the moment denote by T the convex hull of S_0, by T_i the set of U, V channels defined by T in the same way that S_i is defined by S_0, and by T_i^* the convex hull of T_i. It is not hard to verify that

$$S_2^* \supset T_2, \; S_3^* \supset T_3, \text{ so that } S_2^* = T_2^*, \; S_3^* = T_3^*.$$

We conclude that

$$\pi_i(|D|, N, S_0^*) = \pi_i(|D|, N, S_0) \qquad \text{for } i = 2, 3, \tag{1}$$

a fact which will be used later.

We shall call a number $R \geq 0$ an *attainable rate of type* i for S_0 if

$$\pi_i([2^{RN}], N, S_0) \to 0 \text{ as } N \to \infty.$$

The upper bound of the set of attainable rates of type i for S_0 will be called the *type i capacity* of the set S_0 and denoted by $R_i(S_0)$. Thus if R is an attainable rate of type i we can, by random encoding in large blocks, transmit R binary symbols per transmission period, with small error probability.

If, in the definition of π_i above, we had minimized over pure codes instead of random codes, we would have obtained numbers $r_i(S_0)$, which we shall call the *type i pure capacity* of S_0. The present authors in an earlier paper obtained a simple formula for $r_1(S_0)$. The principal result of the present paper is that

$$R_3(S_0) = R_2(S_0) = R_1(S_0^*) = r_1(S_0^*),$$

where S_0^* is the convex hull of S_0. In addition we show that always $R_1(S_0) = r_1(S_0)$ and give an example in which $R_3(S_0) > 0$, $r_2(S_0) = r_3(S_0) = 0$. The evaluation of $r_2(S_0)$ and $r_3(S_0)$ for general S_0 remains unsettled.

We may already conclude from (1) that

$$R_i(S_0) = R_i(S_0^*) \qquad \text{for } i = 2, 3. \tag{2}$$

2. Direct half of principal result. For any random variable X with a finite set of values x, we denote by $I(X)$ the random variable whose value when $X = x$ is $-\log_2 P\{X = x\}$. For any two random variables X, Y, each with a finite set of values, we define

$$\begin{aligned} I(X \mid Y) &= I(X, Y) - I(Y) \\ J(X, Y) &= I(X) + I(Y) - I(X, Y) \\ &= I(X) - I(X \mid Y) \\ &= I(Y) - I(Y \mid X). \end{aligned}$$

$I(X)$ is usually called the information, entropy, or uncertainty in X, $I(X \mid Y)$ the information in X given Y, and $J(X, Y)$ the mutual information in X, Y. These concepts, introduced by Shannon [5], are basic in information theory.

Associated with each probability distribution α on A and A, B channel s is a probability distribution $P_{\alpha s}$ on the set $A \times B$ of pairs (a, b), defined by

$$P_{\alpha s}(a, b) = \alpha(a)s(a, b).$$

Let X, Y be the input, output variables on $A \times B$: $X(a, b) = a$, $Y(a, b) = b$ and define, for any closed subset $S \subset M(A, B)$,

$$H_\alpha(S) = \min_{s \in S} E_{\alpha s} J_{\alpha s}(X, Y),$$

$$H(S) = \max_{\alpha} H_\alpha(S),$$

where the subscripts αs indicate that expectation and mutual information are with respect to $P_{\alpha s}$.

THEOREM 1. *$R_3(S_0) \geqq H(S_0^*)$, where S_0^* is the convex hull of S_0.*

PROOF. We shall first suppose S_0 finite. It suffices to show that, for any α and any number σ with $0 < \sigma < H_\alpha(S_0^*)$, the number $H_1 = H_\alpha(S_0^*) - \sigma$ is an attainable type 3 rate for S_0. Let δ be any number for which $0 < |B|\delta \leqq 1$, where $|B|$ is the number of elements in B, and let s_δ be the B, B channel whose nondiagonal elements are all equal to δ, so that its diagonal elements are all equal to $1 - (|B| - 1)\delta$. Finally, let q be any probability distribution on the finite set F of jamming strategies f.

Let us choose a sequence $X_N^* = (X_1, \cdots, X_N)$ of N independent input variables, each with distribution α, and let L be a jamming strategy, selected independently of X_N^* with distribution q. The input sequence X_N^* and jamming strategy L determine a sequence of output variables $Y_N^* = (Y_1, \cdots, Y_N)$. We use Y_n as an input variable on the B, B channel s_δ, and let Z_n be the resulting output variable. Write $Z_n^* = (Z_1, \cdots, Z_n)$, $n = 1, \cdots, N$. Then

$$
\begin{aligned}
&P\{Y_N^* = v \mid X_N^* = u\} = s(u, v), \qquad s = \sum q(f)s_f, \\
(3)\qquad &P\{Z_n = b \mid X_N^*, Y_N^*, L\} = s_\delta(Y_n, b), \\
&P\{(X_n, Z_n) = (a, b) \mid X_{n-1}^*, Y_{n-1}^*, Z_{n-1}^*, L\} = P_{\alpha s^* s_\delta}(a, b),
\end{aligned}
$$

where s^* is the element of S_0 selected by L for the nth transmission period when the previous input-output history is X_{n-1}^*, Y_{n-1}^*. From (3) we obtain

$$
\begin{aligned}
&P\{(X_n, Z_n) = (a, b) \mid X_{n-1}^*, Z_{n-1}^*\} = \sum_{y,f} P\{Y_{n-1}^* = y, L = f \mid X_{n-1}^*, Z_{n-1}^*\} \\
(4)\qquad &\cdot P\{(X_n, Z_n) = (a, b) \mid X_{n-1}^*, Z_{n-1}^*, Y_{n-1}^* = y, L = f\} = P_{\alpha t s_\delta}(a, b),
\end{aligned}
$$

where $t = t(X_{n-1}^*, Z_{n-1}^*) \varepsilon S_0^*$, the convex hull of S_0.

We shall find an upper bound for $P\{J(X_N^*, Z_N^*) \leqq N(H_1 + \gamma)\}$, where γ is a positive number less than σ. We write

$$J(X_N^*, Z_N^*) = \sum_{n=1}^{N} [J(X_n^*, Z_n^*) - J(X_{n-1}^*, Z_{n-1}^*)] = \sum_{n=1}^{N} J_n,$$

where

$$J_n = I(X_n) + I(Z_n \mid Z_{n-1}^*) - I((X_n, Z_n) \mid (X_{n-1}^*, Z_{n-1}^*)).$$

Let us fix x^*, z^* and denote by μ the conditional joint distribution of (X_n, Z_n) given $X_{n-1}^* = x^*$, $Z_{n-1}^* = z^*$ and by β the conditional distribution of Z_n given $Z_{n-1}^* = z^*$. The conditional distribution of J_n, given $X_{n-1}^* = x^*$, $Z_{n-1}^* = z^*$ is then that of $T = I(X) - \log_2\beta(Z) - I(X, Z)$, where X, Z are the input-output variables on $A \times B$ and μ is the distribution on $A \times B$. Now

$$T = J(X, Z) + \log_2\beta'(Z) - \log_2\beta(Z),$$

where β' is the distribution of Z. Since

$$E(\log_2\beta'(Z) - \log_2\beta(Z)) = -\sum\beta'(z) \log_2(\beta(z)/\beta'(z)) \geq 0$$

(using convexity of $-\log_2$), we obtain $ET \geq EJ(X, Z)$. From (4), μ is a distribution P_{at_0} for some $t \varepsilon S_0^*$, so that, denoting by S_δ^* the set of all A, B channels of the form ts_δ, $t \varepsilon S_0^*$, we have

$$ET \geq H_\alpha(S_\delta^*) = h(\delta). \tag{5}$$

We next find an upper bound for $|T|$. We have $T = -\log_2\beta(Z) - I(Z \mid X)$. Now $\beta(b) \geq \delta$ and $ts_\delta(a, b) \geq \delta$ for all a, b. Thus, since $0 \leq -\log_2\beta(Z) \leq -\log_2\delta$ and $0 \leq I(Z \mid X) \leq -\log_2\delta$, we have

$$|T| \leq -\log_2\delta. \tag{6}$$

Using (5) and (6), we find a bound for the moment generating function of the variable $T_1 = T - h(\delta) + \lambda$, where λ is a positive number. From (5), (6) we obtain $E(T_1) \geq \lambda$, $|T_1| \leq \lambda - \log_2\delta = Q = Q(\lambda, \delta)$. For $t \leq 0$, we have $e^{tT_1} \leq 1 + tT_1 + [(tQ)^2/2]e^{|t|Q}$, so that $\phi(t) = Ee^{tT_1} \leq 1 + \lambda t + [(tQ)^2/2]e^{|t|Q}$. From now on, we restrict λ, δ to the set

$$\lambda/Q \leq \log(4/3). \tag{9}$$

With this restriction, and $t_0 = -\lambda/Q^2$, we obtain

$$\phi(t_0) \leq 1 - (\lambda^2/3Q^2) = \rho_1 = \rho_1(\lambda, \delta). \tag{10}$$

Now ϕ is the conditional moment generating function of $J_n - h(\delta) + \lambda$, given $X_{n-1}^* = x^*$, $Z_{n-1}^* = z^*$. It follows that $E(\exp t_0(\sum_{n=1}^N (J_n - h(\delta) + \lambda))) \leq \rho_1^N$, so that

$$P\{J(X_N^*, Z_N^*) \leq N(h(\delta) - \lambda)\} \leq \rho_1^N(\delta, \gamma). \tag{11}$$

Now $h(\delta) \to H_\alpha(S_0^*)$ as $\delta \to 0$. Choose δ_0 sufficiently small so that

$$h(\delta_0) > H_\alpha(S_0^*) - \sigma + \gamma = H_1 + \lambda$$

and $h(\delta_0) - H_1 - \gamma \leq -\log_2 \delta_0 \log(4/3)$, and set $\lambda_0 = h(\delta_0) - H_1 - \gamma$. From (11) we obtain

$$P\{J(X_N^*, Z_N^*) \leq N(H_1 + \gamma)\} \leq \rho^N = \rho^N(\sigma - \gamma) \tag{12}$$

where $\rho = \rho_1(\lambda_0, \delta_0) < 1$ and depends only on $\sigma - \gamma$ and the modulus of continuity of the function κ. Inequality (12) is the first, and most difficult, step in our proof.

Now

$$P\{Z_N^* = v \mid X_N^* = u\} = \sum_{v'} P\{Y_N^* = v' \mid X_N^* = u\}P\{Z_N^* = v \mid Y_N^* = v'\} = ss_\delta(u, v),$$

where s is the U, V channel defined in (3) and s_δ is the V, V channel which sends inputs Y_N^* into outputs Z_N^*, with $\delta = \delta_0$. We now apply a fundamental inequality of Shannon [6], which asserts the existence, for any message set D

with $|D| \leqq 2^{NH_1}$, of a pure U, V code $c = (s_1, s_2)$, whose average error probability, on channel ss_3, is at most $P\{J(X_N^*, Z_N^*) \leqq N(H_1 + \gamma)\} + 2^{-N\gamma}$. Thus, using (12), we obtain $\pi(ss_3, c) \leqq 2\rho_2^N$, where $\rho_2 = \min_{0<\gamma<\delta} \max(2^{-\gamma}, \rho(\sigma - \gamma)) < 1$. Now $\pi(ss_3, c) = \pi(s, c^*)$, where $c^* = (s_1, s_3s_2)$.

We have now proved the

Lemma. *There is a constant $\rho_2 < 1$ such that, for $|D| = [2^{NH_1}]$ and any probability distribution q on the set F of U, V jamming strategies, there is a D, U, V code c^* for which*

$$\sum_f q(f)\pi(s_f, c^*) \leqq 2\rho_2^N .$$

We now consider the two-person zero sum game in which the pure strategies for Player I are the U, V jamming strategies f, the pure strategies for Player II are the pure D, U, V codes c, and the payoff to Player I for f, c is $\pi(s_f, c)$, the average error probability for code c on the channel sf determined by the jamming strategy f. The lemma asserts that, for any given mixed strategy q of Player I, there is a corresponding strategy for Player II which makes the payoff to I at most $2\rho_2^N$. The minimax theorem then asserts the existence of a mixed strategy for Player II, i.e., a probability distribution k over the set C of pure D, U, V codes, for which $\Sigma k(c)\pi(s_f, c) = \pi(s_f, k) \leqq 2\rho_2^N$ for all jamming strategies f, i.e., $\pi(s, k) \leqq 2\rho_2^N$ for all $s \in S_3$. Thus

$$\pi_3([2^{H_1N}], N, S_0) \leqq 2\rho_2^N \to 0 \text{ as } N \to \infty,$$

H_1 is an admissible rate of type 3, and the proof of Theorem 1 is complete for the case of finite S_0.

The restriction to finite S_0 was made only to avoid irrelevant details, e.g., measurability of jamming strategies. This restriction can now easily be removed by approximation. For an arbitrary S_0, let T be any set which contains S_0 and which is the convex hull of a finite set. Clearly $R_3(S_0) \geqq R_3(T)$, and we have shown that $R_3(T) \geqq H(T)$. Thus $R_3(S_0) \geqq \sup_T H(T)$. It is not difficult to show that $\sup_T H(T) = H(S_0^*)$, completing the proof.

3. Converse half of principal result.

Theorem 2. *For any closed S_0, $R_1(S_0) \leqq H(S_0)$.*

Proof. It was proved in [1] that $r_1(S_0) \leqq H(S_0)$. The present proof is a minor modification of the earlier one. Again, we shall use

Fano's inequality [2], [7]. *For any two random variables W, W',*

$$EI(W \mid W') \leqq -[g \log_2 g + (1 - g) \log_2(1 - g)] + g \log_2(G - 1),$$

where $g = \Pr\{W \neq W'\}$ and G is the number of values of W.

We consider a random (D, U, V) code k, take any U, V channel $s \in S_1$, and suppose that a message is selected from D with a uniform distribution and transmitted over s using k. We denote by W, X, Y, W' the resulting message, U, V input, U, V output, and estimated message respectively. Let $g = \pi(s, k) =$

$\Pr\{W' \neq W\}$. Let us denote by Z the pure code selected, so that Z is independent of W and has distribution k. Then

$$
\begin{aligned}
EJ(X, Y \mid Z) &\geqq EJ(W, W' \mid Z) \\
&= EI(W) - EI(W \mid W', Z) \\
&\geqq EI(W) - EI(W \mid W') \\
&\geqq (1 - g)\log_2 |D| - 1,
\end{aligned} \tag{13}
$$

where the last inequality is obtained from Fano's inequality. Also

$$
\begin{aligned}
EJ(X, Y \mid Z) = EI(Y \mid Z) - EI(Y \mid X, Z) &= EI(Y \mid Z) - EI(Y \mid X) \\
&\leqq EI(Y) - EI(Y \mid X) = EJ(X, Y).
\end{aligned}
$$

Combining this inequality with (13) yields

$$EJ(X, Y) \geqq (1 - g)\log_2 |D| - 1, \tag{14}$$

i.e.,

$$g = \pi(s, k) \geqq 1 - [EJ(X, Y) + 1/\log_2 |D|]. \tag{15}$$

Since the distribution of X is independent of s, we maximize (15) over $s \in S_1$, then minimize over k, to obtain

$$\pi_1(|D|, N, S_0) \geqq 1 - [H(S_1) + 1]/[\log_2 |D|]. \tag{16}$$

But, as shown in [1], $H(S_1) = NH(S_0)$, so that

$$\pi_1([2^{RN}], N, S_0) \geqq 1 - [NH(S_0) + 1]/[\log_2(2^{RN})]. \tag{17}$$

Thus if R is an admissible rate of type 1, $\lim_{N\to\infty} [NH(S_0) + 1]/[\log_2(2^{RN})] \geqq 1$, i.e., $R \leqq H(S_0)$. This completes the proof.

We summarize our results in

THEOREM 3. *For any S_0,*

$$R_3(S_0) = R_2(S_0) = R_1(S_0^*) = r_1(S_0^*),$$

where S_0^* is the convex hull of S_0. Also $R_1(S_0) = r_1(S_0) = H(S_0)$.

PROOF. That $r_1(S_0) = H(S_0)$ was shown in [1]. Since $r_1(S_0) \leqq R_1(S_0)$ and, from Theorem 2, $R_1(S_0) \leqq H(S_0)$, we have $R_1(S_0) = r_1(S_0) = H(S_0)$. The chain of inequalities

$$H(S_0^*) \leqq R_3(S_0) \leqq R_2(S_0) = R_2(S_0^*) \leqq R_1(S_0^*) \leqq H(S_0^*)$$

completes the proof of Theorem 3.

An example and an open question. We have associated with a set S_0 of A, B channels six capacities, according as (a) we face (1) the same unknown channel in S_0 each period, (2) an unknown channel varying arbitrarily in S_0 from period to period, or (3) an unknown channel in S_0, selected each period by a jammer with knowledge of previous inputs and outputs, and (b) we are restricted to

pure codes or are allowed to use random codes. Of these six numbers, we have evaluated four: $r_1(S_0)$ and $R_i(S_0)$, $i = 1, 2, 3$.

The evaluation of $r_2(S_0)$, $r_3(S_0)$ remains unsolved. We conclude with an example in which $r_2(S_0) = r_3(S_0) = 0$, while $R_2(S_0) = R_3(S_0) = \frac{1}{2}$. This example illustrates that, against an unknown arbitrarily varying channel, or against a jammer, random codes are a real improvement over pure codes.

In our example, S_0 consists of two noiseless channels, labeled 0 and 1. Each channel has two inputs, 0 and 1, and three outputs, 0, 1, and 2. Channel i transmits input i perfectly, but changes the other input $1 - i$ into 2:

Input	*Channel* 0 *output*	*Channel* 1 *output*
0	0	2
1	2	1

We shall prove that, for any number N and any pure D, U, V code $c = (s_1, s_2)$, there is a channel $s \in S_2$ for which

$$\pi(s, c) \geqq (G - 1)/2G, \tag{18}$$

where $G = |D|$ = number of messages in D.

Thus no set with two or more messages can be transmitted by a pure code with average error probability less than $\frac{1}{4}$ over every sequence of channels in S_0, no matter how many transmission periods are allowed. It follows that $r_2(S_0) = 0$, and *a fortiori* $r_3(S_0) = 0$. On the other hand, our formula

$$R_3(S_0) = \max_{\alpha} \min_{s \in S_0^*} E_{\alpha s} J_{\alpha s}(X, Y)$$

yields $R_3(S_0) = \frac{1}{2}$, with $\alpha = (\frac{1}{2}, \frac{1}{2})$ as the maximizing input distribution and the channel s with matrix

$$\left\| \begin{matrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{matrix} \right\|:$$

the midpoint of the channels in S_0, as the minimizing channel in S_0^*.

To verify (18), let N be any positive integer, let D be any message set with $|D| = G$ elements, and let $c = (s_1, s_2)$ be any pure D, U, V code. Let x_{dn} denote the nth input specified by c for transmitting message d, and let x_d denote the vector whose coordinates are x_{dn}, $n = 1, 2, \cdots, N$: x_d is the vector in U for which $s_1(d, x_d) = 1$. Let us denote by $s(d)$ that U, V channel in S_2 which transmits x_d perfectly: $s(d)$ has channel number x_{dn} as its nth coordinate. We note that the output v corresponding to any input u and any U, V channel $s \in S_2$ has for its nth coordinate the common vaue of the nth coordinate of u and the number of the nth channel of s, if these numbers agree, and has 2 if they do not. Thus, denoting this output vector by $v(u, s)$, we have

$$v(x_d, s(d')) = v(x_{d'}, s(d)).$$

The probability $p(d, d')$ of an error in transmitting message d over channel

$s(d')$ is 0 if $v(x_d, s(d'))$ is decoded as d, and 1 otherwise. If $d' \neq d$, the vector $v(x_d, s(d')) = v(x_{d'}, s(d))$ cannot be decoded as both d and d', so that $p(d, d') + p(d', d) \geqq 1$ for $d' \neq d$. Summing this inequality over all pairs d, d' with $d' \neq d$ yields

$$2G \sum_{d} \pi(s(d), c) \geqq G(G - 1),$$

so that, for some d, $\pi(s(d), c) \geqq (G - 1),/2G$, and (18) is verified.

REFERENCES

[1] David Blackwell, Leo Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 1229–41.

[2] R. M. Fano, "Statistical theory of communication," notes on a course given at the Massachusetts Institute of Technology, 1952, 1953.

[3] H. W. Kuhn, "Extensive Games and the Problem of Information, Contributions to the Theory of Games," Vol. II, *Ann. Math. Studies*, No. 28, pp. 103–216, Princeton, 1953.

[4] Claude E. Shannon, "A note on a partial ordering for communications channels," *Information and Control* Vol. 1 (1958), pp. 390–398.

[5] C. E. Shannon, "A mathematical theory of communication," Bell System Tech. J., Vol. 27 (1948), pp. 379–423, 623–656.

[6] Claude E. Shannon, "Certain results in coding theory for noisy channels," *Information and Control*, Vol. 1 (1957), pp. 6–25.

[7] A. Feinstein, *Foundation of Information Theory*, McGraw-Hill, New York, 1958, pp. 35–36.

ON THE ESTIMATION OF THE SPECTRUM OF A STATIONARY STOCHASTIC PROCESS

By K. R. Parthasarathy

Indian Statistical Institute, Calcutta

1. Introduction. Recently many authors have been interested in the problem of estimating the spectral density function of a weakly stationary process. Under assumptions of linearity of the process and existence of derivatives of the spectral density, U. Grenander and M. Rosenblatt [1] have investigated the asymptotic behaviour of various estimates. E. Parzen [2] has investigated the asymptotic behaviour of different types of errors of the estimates under assumptions of fourth order stationarity and exponential or algebraic decrease of the covariance sequence.

In this paper, the problem of estimating the spectral distribution as well as the spectral density (if it exists) of a weakly stationary process is solved under the sole assumption that the sample covariances converge almost surely and in mean to the true covariances. The relevance of Bochner's work on Fourier analysis [3], in obtaining more exact expressions for the bias of estimates, is pointed out. The existence of estimates which converge uniformly strongly to the spectral density of the process is proved under the assumption that the density has an absolutely convergent Fourier series. It should be added that only questions of consistency are discussed here and, no attempt is made to derive the asymptotic distribution of the estimates.

2. Estimates of the Spectral Distribution Function.

Definitions: We suppose that $x_1, x_2, \cdots x_N$ are observations at N consecutive time points on a discrete weakly stationary stochastic process

$$\{x_t\}(t = \cdots, -1, 0, 1, \cdots),$$

with the well-known spectral representation (cf. [1])

$$(2.1)\quad x_t = \int_{-\pi}^{\pi} e^{it\lambda}\, dZ(\lambda); \qquad Ex_t = 0; \qquad \rho_\nu = \rho_{-\nu} = Ex_t\, x_{t+\nu} = \int_{-\pi}^{\pi} e^{i\nu\lambda}\, dF(\lambda),$$

where $Z(\lambda)$ is an orthogonal stochastic set function (cf. [1]) and $F(\lambda)$ is a monotonic right continuous function in $[-\pi, \pi]$. It is easily seen that

$$(2.2)\qquad \hat{\rho}_\nu = \hat{\rho}_{-\nu} = (x_1x_{1+|\nu|} + \cdots + x_{N-|\nu|}x_N)/(N - |\nu|)$$

is an unbiased estimate of ρ_ν. We shall consider the following estimate of the spectral distribution:

$$(2.3)\qquad \hat{F}_N(\lambda) = 1/2\pi \sum_{k=-R(N)}^{+R(N)} a_{k,N}\cdot(\hat{\rho}_k/ik)e^{ik\lambda},$$

Received July 20, 1959; revised January 19, 1960.

[1] $\rightarrow$ implies the usual weak convergence of distributions.

where the term corresponding to $k = 0$ is $a_{0,N}(\lambda + \pi)$, and the $a_{k,N}$ are constants chosen such that the following conditions are satisfied:

1) $a_{k,N} \to 1$ as $N \to \infty$ for each fixed k,

2) $a_{k,N} = a_{-k,N}$,

3) $\hat{F}_N(\lambda)$ is a distribution function.

As is known from previous work [1], [2], it is advantageous to choose $R(N) = o(N)$. We shall now state, without proof, a theorem concerning the convergence of the estimates $\hat{F}_N(\lambda)$.

THEOREM 2.1: *If $\{x_i\}$ is a weakly stationary process with a spectral distribution $F(\lambda)$, and the sample covariances converge almost surely to the true convariances, then*[1] $P[\hat{F}_N \to F] = 1$. *If, however, $F(\lambda)$ is continuous, then*

$$P[\sup_{|\lambda| \leq 2\pi} |\hat{F}_N(\lambda) - F(\lambda)| \to 0 \text{ as } N \to \infty] = 1.$$

If, further, the sample covariances converge in mean to the true covariances, then

$$\lim_{N\to\infty} E \sup_{|\lambda| \leq 2\pi} |\hat{F}_N(\lambda) - F(\lambda)| = 0.$$

The first part of the theorem is contained in Doob [4]; the second part follows by an application of a theorem of Pólya to the effect that the weak convergence of a sequence of distributions to a continuous distribution implies uniform convergence; the last part follows from an easy computation.

The choice of the constants $a_{k,N}$: Our main object is to make a suitable choice of the constants $a_{k,N}$, and to examine the order of the bias, convergence, etc., of the estimates thus obtained. The method we use for this purpose is simply a Fourier analysis. It is based almost entirely on the work of Bochner [3]. We now state the main result of Bochner, in the form required here.

Let $f(x)$ be a continuous periodic function with period 2π and let

$$K(t) = \left(\frac{\sin t/2}{t/2}\right)^2, \qquad M(r) = \int_{-\infty}^{+\infty} \left(\frac{\sin t/2}{t/2}\right)^{2r} dt,$$

$$K_r(t) = \frac{[K(t)]^r}{M(r)},$$

$$S_R^r(x) = \int_{-\infty}^{+\infty} f(x + t)\, R/r\, K_r(Rt/r)\, dt.$$

THEOREM (BOCHNER): *For any continuous periodic function, $f(x)$,*

$$|S_R^r(x) - f(x)| = O[w(4r/R) + 4^{-r}],$$

where

$$w(x) = \max_{|x_1 - x_2| < x} |f(x_1) - f(x_2)|.$$

Write

$$(2.4)\qquad f_N^*(\lambda) = \frac{1}{2\pi N}\int_{-\infty}^{+\infty}\left|\sum_{t=1}^{N} x_t e^{it(\lambda+u)}\right|^2 \frac{R}{r} K_r\left(\frac{Ru}{r}\right) du$$

and

$$(2.5)\qquad F_N^*(\lambda) = \int_{-\pi}^{\lambda} f_N^*(\lambda)\, d\lambda.$$

Then it is possible to write $F_N^*(\lambda)$ as given in (2.3) and to show that all the required conditions are satisfied. Thus, by Theorem 2.1, F_N^* is a consistent estimator of F under very mild conditions. We now state a theorem concerning the bias of F_N^* as an estimate of the spectral distribution.

THEOREM 2.2. *For a weakly stationary process, $\{x_t\}$, with a continuous spectral distribution, F, we have*

$$\sup_{|\lambda|<\pi} | EF_N^*(\lambda) - F(\lambda) | = O[w(4r/R) + 4^{-r} + R/Nw(R^{-1})],$$

where

$$w(x) = \max_{|\lambda_1-\lambda_2|<x} | G(\lambda_1) - G(\lambda_2) |, \qquad x \leqq 2\pi,$$

and

$$G(\lambda) = F(\lambda) - [(\lambda + \pi)/2\pi]\rho_0 .$$

Since the above is an easy consequence of Bochner's theorem, the proof is omitted.

COROLLARY: *If $F(\lambda)$ satisfies Lipschitz's condition, i.e.*

$$| F(\lambda_1) - F(\lambda_2) | < c | \lambda_1 - \lambda_2 | ,$$

where c is a constant, then $w(x) < cx$ for any $x > 0$, and hence

$$\sup | EF_N^*(\lambda) - F(\lambda) | = O[r/R + 4^{-r} + ((R)^{\frac{1}{2}}/N)].$$

Thus, in order to obtain an asymptotically unbiased and consistent estimator of F, we have only to choose r and R such that $r \to \infty$, $R \to \infty$, $r/R \to 0$ and $R/N \to 0$ as $N \to \infty$ in $F_N^*(\lambda)$.

For Gaussian processes the following theorem can be easily proved.

THEOREM 2.3. *For a Gaussian process with a square integrable spectral density we have*

$$E \sup_{|\lambda| \leqq 2\pi} | F_N^*(\lambda) - F(\lambda) | = O[(\log R/(N)^{\frac{1}{2}}) + w(4r/R) + 4^{-r}].$$

3. Convergence of the Spectral Density. In this section we shall discuss the choice of r and R so that the estimate $f_N^*(\lambda)$ given in (2.4) converges (almost surely) uniformly to the spectral density of the process. Our choice will be such that r and R are not only functions of N but of the observations themselves. It should be noted that, even if r and R depend on the observations, Theorem 2.1 remains valid provided that r and R diverge to infinity with probability one.

We require the following

LEMMA 3.1. *For any weakly stationary process x_t, if $\sum_1^N x_t^2/N$ is convergent with probability one as $N \to \infty$, then, for $\epsilon > 0$,*

$$P[\sup_N \sup_{0 \leq k \leq N^{1-\epsilon}} |\hat{\rho}_k| < \infty] = 1,$$

where $\hat{\rho}_k$ is as in (2.2).

PROOF:

$$|\hat{\rho}_k| = (|\sum_{t=1}^{N-k} x_t x_{t+k}|)/(N-k) \leq 1/(N-k)[(\sum_1^{N-k} x_t^2)(\sum_{k+1}^{N} x_t^2)]^{\frac{1}{2}},$$

so that

$$\sup_{0 \leq k \leq N^{1-\epsilon}} |\hat{\rho}_k| \leq 1/(N - N^{1-\epsilon}) \sum_1^N x_t^2 < [(\sum_1^N x_t^2)/N(1 - 2^{1-\epsilon})] \tag{3.1}$$

for $N \geqq 2$. Since by assumption $(\sum_1^N x_t^2)/N$ converges, the expression on the right side of 3.1 is bounded with probability one. This completes the proof.

Our estimate of the spectral density function is

$$f_N^*(\lambda) = 1/2\pi N \int_{-\infty}^{+\infty} \left| \sum_{t=1}^{N} x_t e^{it(\lambda+u)} \right|^2 R/r\, K_r\,(Ru/r)\, du,$$

which can also be rewritten as

$$f_N^*(\lambda) = 1/2\pi \sum_{m=-R}^{+R} \varphi^r(rm/R)(1 - |m|/N)\hat{\rho}_m e^{im\lambda}, \tag{3.2}$$

where

$$\varphi^r(t) = \int e^{itx} K_r(x)\, dx.$$

We now prove the following

THEOREM 3.1. *Let $\{x_t\}$ be a weakly stationary process, with spectral density function $f(\lambda)$ and covariance sequence $\{\rho_k\}$, which has the property that $\sum_{-\infty}^{+\infty} |\rho_k|$ is convergent. Suppose, further, that the sample variance and covariances converge almost surely, and in the L_1 mean, to the true variance and covariances respectively. Then there exist $R(N, x_1, x_2 \cdots x_N)$ and $r(N, x_1, x_2, \cdots x_N)$ such that*

$$\sup_{|\lambda| < \pi} |f_N^*(\lambda) - f(\lambda)| \to 0$$

almost surely as $N \to \infty$.

PROOF:

$$\begin{aligned} f_N^*(\lambda) &= 1/2\pi \sum_{-R}^{+R} \varphi^r(rm/R)(1 - (|m|/N))(\hat{\rho}_m - \rho_m)e^{im\lambda} \\ &+ 1/2\pi \sum_{-R}^{+R} \varphi^r(rm/R)(1 - (|m|/N))\rho_m e^{im\lambda} = S_1 + S_2, \quad \text{say.} \end{aligned} \tag{3.3}$$

For S_1 we have

$$\sup_\lambda | S_1 | \leqq 1/2\pi \sum_{-R}^{+R} | \hat{\rho}_m - \rho_m | \leqq 1/\pi \sum_{0}^{R} | \hat{\rho}_m - \rho_m | \tag{3.4}$$
$$\leqq 1/\pi R^{1+\delta} \sum_{1}^{[N^{1-\epsilon}]} [(| \hat{\rho}_m - \rho_m |)/m^{1+\delta}]$$

if $R < [N^{1-\epsilon}]$, $\epsilon > 0$, $\delta > 0$. Since for each m, $| \hat{\rho}_m - \rho_m | \to 0$ with probability one, and by Lemma 3.1, $\sum_1^{[N^{1-\epsilon}]} [(| \hat{\rho}_m - \rho_m |)/m^{1-\delta/2}]$ is bounded, we get by Toeplitz's lemma [5] the following:

$$P[\lim_{N\to\infty} \sum_{1}^{[N^{1-\epsilon}]} [(| \hat{\rho}_m - \rho_m |)/m^{1+\delta/2}]\cdot 1/m^{\delta/2} = 0] = 1. \tag{3.5}$$

We choose R such that $R \to \infty$, with probability one, $R < [N^{1-\epsilon}]$, and

$$R = o\left[\sum_{1}^{[N^{1-\epsilon}]} [(| \hat{\rho}_m - \rho_m |)/m^{1+\delta}] \right]^{-1/(1+\delta)}. \tag{3.6}$$

Then

$$P[\sup_\lambda | S_1 | \to 0 \quad \text{as} \quad N \to \infty] = 1.$$

Turning to S_2, we have

$$S_2(\lambda) - f(\lambda) = \int_{-\infty}^{+\infty} [f_N(\lambda + t) - f(\lambda)] R/r\, K_r\, (Rt/r)\, dt, \tag{3.7}$$

where

$$f_N(\lambda) = 1/2\pi \sum_{-N}^{+N} (1 - (| m |/N))\rho_m e^{im\lambda} \tag{3.8}$$

is the Nth Fejer mean of $f(\lambda)$. Since $\sum_{-\infty}^{+\infty} | \rho_k |$ is convergent, $f(\lambda)$ is bounded and continuous. Since $f(\lambda)$ is symmetric in λ, $f(\pi) = f(-\pi)$. Hence, by Fejer's theorem,

$$\lim_{N\to\infty} \sup_{|\lambda| \leqq \pi} | f_N(\lambda) - f(\lambda) | = 0. \tag{3.9}$$

From (3.7) we have

$$\sup_\lambda | S_2(\lambda) - f(\lambda) | \leqq \sup_\lambda \int_{-\infty}^{+\infty} | f_N(\lambda + t) - f(\lambda + t) | \tag{3.10}$$
$$\cdot (R/r) K_r(Rt/r)\, dt + \sup_\lambda \left| \int_{-\infty}^{+\infty} [f(\lambda + t) - f(\lambda)]\, (R/r) K_r(Rt/r)\, dt \right|.$$

Since $\int_{-\infty}^{+\infty} (R/r)K_r(Rt/r)\, dt = 1$, the first term on the right of (3.10) goes to zero as $N \to \infty$. If we choose r such that $r \to \infty$ and $(r/R) \to 0$ as $N \to \infty$, it is easily seen from Bochner's theorem, that the second term also goes to zero with probability one.

We remark that, if we choose $rR = o(N)$ and $r = o(R)$, the theorems of U. Grenander and M. Rosenblatt [1] on the consistency of the spectral estimates for linear processes become applicable.

Finally, let us consider the behaviour of the periodogram of a stationary Gaussian process. It is well-known that the periodogram does not converge to any random variable as the sample size increases to infinity. However, the following theorem holds.

THEOREM 3.2. *For a stationary Gaussian process with a spectral density* $f(\lambda)$ *satisfying Lipschitz's condition,*

$$P\left[\limsup_{N\to\infty}\frac{1}{2\pi}\cdot\frac{\left|\sum_{t=1}^{N}x_t\cos t\lambda\right|^2}{2N\log\log N}+\limsup_{N\to\infty}\frac{1}{2\pi}\frac{\left|\sum_{t=1}^{N}x_t\sin t\lambda\right|^2}{2N\log\log N}=f(\lambda)\right]=1.$$

The proof follows from the analyses of W. Feller [6] and G. Maruyama [7].

REFERENCES

[1] U. GRENANDER AND M. ROSENBLATT, *Statistical Analysis of Stationary Time Series*, John Wiley and Sons, New York, 1957.
[2] E. PARZEN, "On asymptotically efficient consistent estimates of the spectral density function of a stationary time series," *J. Roy. Stat. Soc., (Ser. B)*, Vol. 20 (1958), pp. 303–322.
[3] S. BOCHNER, "Localization of best approximation," in ZYGMUND et al., *Contributions to Fourier Analysis, Ann. Math. Studies*, No. 25, Princeton, 1950.
[4] J. L. DOOB, *Stochastic Processes*, John Wiley and Sons, New York, 1953.
[5] M. LOÈVE, *Probability Theory*, D. Van Nostrand, New York, 1955.
[6] W. FELLER, "The law of the iterated logarithm for identically distributed random variables," *Ann. Math.*, Vol. 47 (1946), pp. 631–638.
[7] G. MARUYAMA, "Harmonic analysis of stationary stochastic processes," *Memoirs of Kyusyu University*, Vol. 4 (1949), pp. 45–106.

EXPECTATIONS OF FUNCTIONALS ON A STOCHASTIC PROCESS

BY EDWARD NELSON AND DALE VARBERG

University of North Dakota and Hamline University

1. Introduction. Let $\{x(t), 0 \leqq t < \infty\}$ be a separable stochastic process with stationary, independent increments, for which $x(0) = 0$ and whose characteristic function is

$$E\{e^{i\xi x(t)}\} = e^{at(\cos\xi - 1)}, \qquad a > 0.$$

One may verify that, if $0 \leqq t_1 < t_2 < \cdots < t_k < \infty$ and m_j is an integer,

$$P\{x(t_k) = m_k, x(t_{k-1}) = m_{k-1}, \cdots, x(t_1) = m_1\}$$
$$= e^{-a(t_k - t_{k-1}) - \cdots - a(t_2 - t_1) - at_1} I_{m_k - m_{k-1}}[a(t_k - t_{k-1}] \cdots I_{m_2 - m_1}[a(t_2 - t_1] I_{m_1}[(at_1)],$$

where $I_n(x) = i^{-n} \cdot J_n(ix)$, $J_n(x)$ being the Bessel function of the first kind. By separability the sample functions, $x(t)$, of this process are simple functions which assume integral values on intervals. They may be interpreted as the monetary gain in coin tossing at random times. To be more precise, $x(t)$ is the sum of a random number, $N(t)$, of independent, identically distributed Bernoulli variables with distribution $P\{x = -1\} = P\{x = 1\} = \frac{1}{2}$, where $N(t)$ is the sample function of a Poisson process ([1], page 398). This process is important in the theory of collective risk and has been studied by Täcklind [2]. Certain similarities between it and the Wiener process led us to attempt to find the expected value of some functionals on this process using a method developed by Kac ([3], Section 3). The principal result of this paper is the following theorem.

THEOREM. *Let*

$$\Psi_n = \int_0^\infty e^{-st} E\left\{\exp\left[-u \int_0^t V(x(\tau))\, d\tau\right], x(t) = n\right\} dt \tag{1.1}$$

where V is non negative. Then Ψ_n satisfies the difference system

$$\begin{gathered} \Psi_{n+1} - (2/a)(s + a + uV_n)\Psi_n + \Psi_{n-1} = -(2/a)\delta_{n,0}, \\ \Psi_n \to 0 \quad \text{as} \quad n \to \pm\infty, \end{gathered} \tag{1.2}$$

where V_n is the value of the function V when $x = n$. (Note: For any function K, $E\{K(x), x(t) = n\}$ means $E\{K(x)\chi(x)\}$ where $\chi(x) = 1$ if $x(t) = n$ and $\chi(x) = 0$ otherwise.)

In Section 2 we outline the proof of the theorem and in Section 3 we give some illustrative examples.

2. Proof of Theorem. In order that we may easily interchange the order of certain limits, we assume first that V is bounded. This restriction will be removed later in the proof. Following the method and notation of Kac we define inductively

Received March 30, 1959; revised January 29, 1960.

$$(2.1) \qquad Q_k(n, t) = \int_0^t \sum_{m=-\infty}^{\infty} V(m)e^{-a(t-\tau)} I_{n-m}[a(t-\tau)]Q_{k-1}(m, \tau)\, d\tau,$$

where $Q_0(n, t) = e^{-at}I_n(at)$. This gives

$$\begin{aligned} Q_k(n, t) = &\int_0^t \int_0^{\tau_k} \cdots \int_0^{\tau_2} \sum_{m_1=-\infty}^{\infty} \cdots \sum_{m_k=-\infty}^{\infty} V(m_1) \cdots V(m_k) \\ &\cdot \exp\,[-a(t-\tau_k) - a(\tau_k - \tau_{k-1}) - \cdots - a(\tau_2 - \tau_1) - a\tau_1] \\ &\cdot I_{n-m_k}[a(t-\tau_k)] \cdot I_{m_k - m_{k-1}}[a(\tau_k - \tau_{k-1})] \cdots I_{m_2-m_1}[a(\tau_2-\tau_1)]I_{m_1}(a\tau_1)\, d\tau_1 \cdots d\tau_k \\ = &\int_0^t \int_0^{\tau_k} \cdots \int_0^{\tau_2} E\{V[x(\tau_1)]V[x(t_2)] \cdots V[x(\tau_k)], x(t) = n\}\, d\tau_1 \cdots d\tau_k \\ = &\,E\left\{\int_0^t \int_0^{\tau_k} \cdots \int_0^{\tau_2} V[x(\tau_1)]V[x(\tau_2)] \cdots V[x(\tau_k)]\, d\tau_1 \cdots d\tau_k, x(t) = n\right\}. \end{aligned}$$

Thus

$$(2.2) \quad Q_k(n, t) = E\left\{\frac{1}{k!}\left(\int_0^t V[x(\tau)]\, d\tau\right)^k, x(t) = n\right\} \leqq \frac{1}{k!} t^k M^k P\{x(t) = n\},$$

where M is an upper bound for V. We define

$$(2.3) \qquad Q(n, t, u) = \sum_{k=0}^{\infty} (-1)^k u^k Q_k(n, t).$$

Using (2.2), we obtain

$$(2.4) \qquad Q(n, t, u) = E\left\{\exp\left[-u \int_0^t V(x(\tau))\, d\tau\right], x(t) = n\right\}.$$

We see immediately that

$$(2.5) \qquad Q(n, t, u) \leqq P\{x(t) = n\} = e^{-at}I_n(at).$$

From (2.1) and (2.3), we find

$$(2.6) \qquad \begin{aligned} &Q(n, t, u) - Q_0(n, t) \\ &\qquad = -u \sum_{m=-\infty}^{\infty} \int_0^t V(m)e^{-a(t-\tau)} I_{n-m}[a(t-\tau)]Q(m, \tau, u)\, d\tau. \end{aligned}$$

Now let, [see (1.1)], $\Psi_n = \int_0^\infty e^{-st}Q(n, t, u)\, dt$, and take the Laplace Transform of both sides in (2.6). This gives (see [4] page 131, Formula 6)

$$(2.7) \qquad \psi_n = \frac{A^{|n|}}{c} - \frac{u}{c} \sum_{m=-\infty}^{\infty} A^{|n-m|}\psi_m V_m,$$

where $c = (s^2 + 2as)^{\frac{1}{2}}$ and $A = a/(s + a + c)$. From (2.7) it can be shown that, for $n \neq 0$,

$$\Psi_{n+1} + \Psi_{n-1} = [(A + A^{-1}) + (u/c)V_n(A^{-1} - A)]\Psi_n,$$

with a similar formula for $n = 0$. The difference system (1.2) now follows easily, the boundary conditions coming from the estimate in (2.5).

Now suppose that $V(x)$ is unbounded and define $V_M(x)$ as $V(x)$ if $V(x) \leqq M$ and 0 otherwise. We have then from (1.2) the difference equation

$$\Psi_{M,n+1} - \frac{2}{a}(s + a + uV_{M,n})\Psi_{M,n} + \Psi_{M,n-1} = -\frac{2}{a}\delta_{n,0}, \tag{2.8}$$

where

$$\Psi_{M,n} = \int_0^\infty e^{-st}E\left\{\exp\left[-u\int_0^t V_M(x(\tau))\,d\tau\right], x(t) = n\right\} dt.$$

By bounded convergence $\lim_{M\to\infty} \Psi_{M,n} = \Psi_n$. Thus, taking limits on both sides of (2.8), we obtain the desired result.

3. Examples.

(a) Let $V(x) = 0$ if $-p < x < q$ and 1 otherwise where p and q are positive integers. We define $\Psi_n^* = \lim_{u\to\infty} \Psi_n$ and note that

$$\Psi_n^* = \int_0^\infty e^{-st}P\{-p < x(\tau) < q \text{ for } 0 \leqq \tau \leqq t, x(t) = n\}\,dt.$$

We observe that, for $-p < n < q$, Ψ_n^* satisfies the difference equation in (1.2) corresponding to this V; hence,

$$\Psi_n^* = \begin{cases} 0 & n \geqq q \\ D_1A^n + D_2A^{-n} & 0 \leqq n \leqq q \\ E_1A^n + E_2A^{-n} & -p \leqq n \leqq 0 \\ 0 & n \leqq -p \end{cases} \tag{3.1}$$

where D_1, D_2, E_1, and E_2 are suitable constants, and

$$A = a/[s + a + (s^2 + 2as)^{\frac{1}{2}}].$$

Let

$$\Psi = \sum_{n=-\infty}^{+\infty} \Psi_n^* = \int_0^\infty e^{-st}P\{-p < x(\tau) < q \text{ for } 0 \leqq \tau \leqq t\}\,dt.$$

Using (3.1), we obtain

$$\Psi = 1/s\cdot(1 - A^p)(1 - A^q)/(1 - A^{p+q}). \tag{3.2}$$

In the special case where $p = \infty$, (3.2) is easily inverted giving

$$P\{\sup_{0\leqq\tau\leqq t} x(\tau) < q\} = 1 - q\int_0^t e^{-a\tau}(I_q(a\tau)/\tau)\,d\tau,$$

a result obtained by Baxter and Donsker in ([5], Section 4).

(b) Let $V(x) = x^2$. The difference equation in (1.2) then becomes

$$\Psi_{n+1} - \frac{2}{a}(s + a + un^2)\Psi_n + \Psi_{n-1} = -\frac{2}{a}\cdot\delta_{n,0}.$$

We define $\Psi(\xi) = \sum_{n=-\infty}^{\infty} e^{2in\xi}\Psi_n = \sum_{n=-\infty}^{\infty} \Psi_n \cos 2n\xi$. Then $\Psi(\xi)$ satisfies the differential system

$$\text{(3.3)}\qquad \begin{aligned} &\Psi''(\xi) - [(4/u)(s + a) - (4a/u)\cos 2\xi]\Psi(\xi) = -4/u, \\ &\qquad\qquad \Psi'(0) = \Psi'(\pi/2) = 0. \end{aligned}$$

To solve (3.3) we consider the differential equation

$$\text{(3.4)}\qquad \Psi''(\xi) + [\mu - (4/u)(s + a) + (4a/u)\cos 2\xi]\Psi(\xi) = 0$$

with the same boundary conditions as in (3.3). The Green's function $G(\xi, \eta)$ for (3.4) is given by

$$\text{(3.5)}\qquad G(\xi, \eta) = \sum_{k=0}^{\infty} \phi_k(\xi)\phi_k(\eta)/\mu_k$$

where μ_k and $\phi_k(\xi)$ are the eigenvalues and normalized eigenfunctions of (3.4) respectively. By Mercer's Theorem ([6], p. 138) the convergence is uniform in ξ and η, the μ_k's all being positive (at least for large s). The solution for $\Psi(\xi)$ in (3.3) is thus given by

$$\text{(3.6)}\qquad \Psi(\xi) = (4/u)\int_0^{\pi/2} G(\xi, \eta)\, d\eta = (4/u)\sum_{k=0}^{\infty} \frac{\phi_k(\xi)}{\mu_k}\int_0^{\pi/2} \phi_k(\eta)\, d\eta.$$

On the other hand, if we let $\lambda = \mu - (4/u)(s + a)$, (3.4) is seen to be Mathieu's equation. Using the notation of ([4], p. 46), we find that

$$\phi_k(\xi) = b_k ce_{2k}(\xi) = b_k \sum_{n=0}^{\infty} A_{2k,2n}\cos 2n\xi$$

where $b_k = (2/\pi)^{\frac{1}{2}}$ if $k = 0$ and $b_k = 2/(\pi)^{\frac{1}{2}}$ if $k \neq 0$.
Upon substituting in (3.6), we obtain

$$\begin{aligned} \Psi(\xi) &= 2\pi/u \sum_{k=0}^{\infty} b_k^2 A_{2k,0}/\mu_k \sum_{n=0}^{\infty} A_{2k,2n}\cos 2n\xi \\ &= (4/u)\sum_{k=0}^{\infty} A_{2k,0}/\mu_k \sum_{n=-\infty}^{+\infty} A_{2k,2n}\cos 2n\xi \end{aligned}$$

where $A_{2k,2n} = A_{2k,-2n}$ for $n < 0$. After interchanging the order of summation, we have

$$\Psi(\xi) = (4/u)\sum_{n=-\infty}^{\infty}\sum_{k=0}^{\infty} \{A_{2k,0}A_{2k,2n}/[\lambda_k + (4/u)(s + a)]\}\cos 2n\xi.$$

By the uniqueness of the Fourier coefficients, it follows that

$$\Psi_n = (4/u) \sum_{k=0}^{\infty} A_{2k,0} A_{2k,2n} / [\lambda_k + (4/u)(s + a)].$$

Inverting with respect to s, we obtain

$$E\left\{\exp\left[-u \int_0^t [x(\tau)]^2 \, d\tau\right], x(t) = n\right\} = \sum_{k=0}^{\infty} A_{2k,0} A_{2k,2n} e^{-(a+u\lambda_k/4)t}.$$

REFERENCES

1. J. L. Doob, *Stochastic Processes*, John Wiley and Sons, New York, 1953.
2. Sven Täcklind, "Sur le risque de ruine dans des jeux inéquitables," *Skand. Aktuarietids.*, Vol. 25, (1942), pp. 1–42.
3. M. Kac, "On some connections between probability theory and differential and integral equations," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 189–215.
4. Wilhelm Magnus and Fritz Oberhettinger, *Formulas and Theorems for the Special Functions of Mathematical Physics*, Chelsea, New York, 1949.
5. Glen Baxter and M. D. Donsker, "On the distribution of the supremum functional for processes with stationary independent increments," *Trans. Amer. Math. Soc.*, Vol. 85, (1957), pp. 73–87.
6. R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Vol. 1, Interscience, New York, 1953.

ERGODOCITY OF QUEUES IN SERIES[1]

By J. Sacks

Columbia University

1. Introduction. We are interested here in determining when a queueing system consisting of several queues in series is ergodic. To define what is meant by queues in series let us consider the case where there are two servers. The definition of two queues in series is given as follows: The nth individual arriving to the queueing system enters, at his time of arrival, a queue (queue 1) in front of the first server. He waits there until all the individuals in front of him have been served in the first server at which time he begins his service. Upon completion of his service the nth individual enters a queue (queue 2) in front of the second server, waits there until all the individuals in front of him have completed their service in the second server, and at that time he begins his own service. Queue 1 and queue 2 are now said to be in series. Putting matters more concisely we can say that two queues are in series if the output of the first queue is the input of the second queue.

To define what we mean by the ergodicity of this queueing system, let W_n be the waiting time in queue 1 of the nth individual and let W_n^* denote his waiting time in queue 2. The queueing system is said to be ergodic if the joint distribution of (W_n, W_n^*) converges, as $n \to \infty$, to a probability distribution. Assuming existence of first moments for the two service time random variables and the interarrival random variable (the sequence of interarrival time random variables is assumed to be a sequence of independent and identically distributed random variables and the same is assumed for each of the two sequences of service time random variables) we are able, in Theorems 1 and 2 below, to characterize when $\{(W_n, W_n^*)\}$ has a limiting probability distribution.

The method we use to characterize ergodicity is first to show (Lemma 2 below) that the distribution function of (W_n, W_n^*) converges to a limit as $n \to \infty$ though the limit may not be a probability distribution function. This is the easy part of the argument. The second part of the argument is to show under appropriate conditions (see Theorem 1) that (W_n, W_n^*) is bounded in probability so that as $n \to \infty$ no probability escapes to infinity and this yields the fact that the limit shown to exist in the first part is a bonafide probability distribution. The last part of the characterization lies in showing that when the conditions for Theorem 1 are not satisfied then either W_n or W_n^* goes to $+\infty$ in probability (Theorem 2). This outline of the argument is quite the same as the outline of the argument used by Kiefer and Wolfowitz [3] for the queueing system they consider. The details of the first part are strongly related to those in [3]. The

---

Received August 14, 1959; revised March 28, 1960.

[1] This research was sponsored by the Office of Naval Research under Contract Number Nonr-266(33), Project Number 042-034. Reproduction in whole or in part is permitted for any purpose of the United States Government.

means whereby the second and third parts of the argument are accomplished depends on knowledge of the behavior of the maximum of partial sums of independent and identically distributed random variables. This was first utilized by Lindley [4] in his treatment of the one server queueing system.

Burke [2], Reich [5] and others (see [2] and [5] for other references) have considered queues in series when the service time random variables and interarrival random variables are exponentially distributed. Akaike [1] considers a problem of ergodic behavior of queues in series related to the one we treat here. Akaike assumes that all the random variables in sight take on values which are integral multiples of some fixed positive number so that the waiting time process is a discrete process (our random variables have no such restriction). Furthermore he assumes that the nth customer cannot enter the second queue before customer $n + 1$ arrives at the first queue; thus if customer n finishes service at server 1 before $n + 1$ arrives, customer n must wait until customer $n + 1$ arrives before entering the second queue. Our assumptions are that the nth customer enters the second queue immediately after finishing service at server 1.

We have only talked about the case of two servers which gives rise to two queues in series. It is simple to see how to define s queues in series when there are s servers—this giving rise to s different waiting times to worry about. All our previous remarks for the two-server case are valid for the s-server case. We have separated the treatment of the two-server case from that of the s-server case in order to avoid confusing notational problems with the principal ideas.

2. The Two-Server Case. In this section we will consider the case where there are two queues in series, the output of the first queue being the input to the second queue.

Let τ_n be the time at which the nth individual enters the system. Let R_n denote the service time in the first server of individual n and let ρ_n be the service time of individual n in the second server. Let $g_{n+1} = \tau_{n+1} - \tau_n$. We assume that each of the three sequences $\{R_n ; n \geqq 1\}$, $\{g_n ; n \geqq 2\}$, $\{\rho_n ; n \geqq 1\}$ is a sequence of independent and identically distributed random variables and that the three sequences are mutually independent. Furthermore we assume that the R_n's, g_n's, and ρ_n's are non-negative random variables and that $ER_1 < \infty$, $Eg_2 < \infty$, $E\rho_1 < \infty$.

Let W_n be the waiting time in the first queue of the nth person and let W_n^* be the waiting time in the second queue of the nth person. The waiting time, of course, is the time between arrival at the queue and the beginning of service. To establish a relationship between (W_n, W_n^*) and (W_{n+1}, W_{n+1}^*) observe that the nth individual leaves the first server (enters the second queue) at time $\tau_n + W_n + R_n$ and leaves the second server at time $\tau_n + W_n + R_n + W_n^* + \rho_n$, while the $(n + 1)$th individual arrives at the first queue at time τ_{n+1} and at the second queue at time $\tau_{n+1} + W_{n+1} + R_{n+1}$. Thus the $(n + 1)$th person waits 0 time in the first queue if $\tau_n + W_n + R_n \leqq \tau_{n+1}$ i.e., if $W_n + R_n - g_{n+1} \leqq 0$, and waits $W_n + R_n - g_{n+1}$ if the last quantity is positive. Stated

more concisely

$$W_{n+1} = \max [0, W_n + R_n - g_{n+1}]. \tag{2.1}$$

Similarly,

$$W^*_{n+1} = \max [0, W^*_n + \rho_n - R_{n+1} + R_n - g_{n+1} + W_n - W_{n+1}]. \tag{2.2}$$

(2.1) and (2.2) are valid for all $n \geqq 1$ with $W_1 = W^*_1 = 0$.

Let $Z_n = (W_n, W^*_n)$. $\{Z_n\}$ is not a Markov process but putting $Y_n = (Z_n, R_n)$ provides us with a sequence $\{Y_n\}$ which is a Markov process with stationary transition probabilities. These considerations will enable us to prove Lemma 2 below which is the first step in characterizing when $\{Z_n\}$ is an ergodic process.

Let $t = (t_1, t_2)$, $x = (x_1, x_2)$ with t_1, t_2, x_1, x_2 all nonnegative numbers.

LEMMA 1: *$P\{Z_n \leqq t \mid Z_1 = x, R_1 = r\} \leqq P\{Z_n \leqq t \mid Z_1 = 0, R_1 = r\}$ for all n, x, t, r.*

PROOF: Fix a point ω in the sample space of $R_2, \cdots, R_n, g_2, \cdots, g_n, \rho_1, \cdots \rho_{n-1}$, and let

$$W_1(\omega, x) = x_1, \qquad W^*_1(\omega, x) = x_2$$

$$W_j(\omega, x) = \max [0, W_{j-1}(\omega, x) + R_{j-1}(\omega) - g_j(\omega)]$$

$$W^*_j(\omega, x) = \max [0, W^*_{j-1}(\omega, x) + \rho_{j-1}(\omega) - R_j(\omega) + R_{j-1}(\omega) - g_j(\omega) + W_{j-1}(\omega, x) - W_j(\omega, x)],$$

for $2 \leqq j \leqq n$. It is clear that $W_j(\omega, 0) \leqq W_j(\omega, x)$ for each j. Observing that

$$\begin{aligned} R_{j-1}(\omega) + W_{j-1}(\omega, x) - W_j(\omega, x) &= R_{j-1}(\omega) + W_{j-1}(\omega, x) - \max [0, W_{j-1}(\omega, x) + R_{j-1}(\omega) - g_j(\omega)] \\ &= \min [W_{j-1}(\omega, x) + R_{j-1}(\omega), g_j(\omega)], \end{aligned}$$

we have

$$W^*_j(\omega, x) = \max [0, W^*_{j-1}(\omega, x) + \rho_{j-1}(\omega) - R_j(\omega) - g_j(\omega) + \min [W_{j-1}(\omega, x) + R_{j-1}(\omega), g_j(\omega)]]$$

and it follows easily that $W^*_j(\omega, 0) \leqq W^*_j(\omega, x)$ for all $2 \leqq j \leqq n$. Lemma 1 is now seen to be true.

LEMMA 2: *$P\{Z_n \leqq t \mid Z_1 = 0\} \to F(t)$ as $n \to \infty$ where F is a two-dimensional distribution function whose variation over two-dimensional space may be less than one i.e., F may not be a probability distribution function.*

PROOF: Let $H(x, r) = P\{Z_2 \leqq x, R_2 \leqq r \mid Z_1 = 0\}$. Then

$$\begin{aligned} &P\{Z_{n+1} \leqq t \mid Z_1 = 0\} \\ &\quad = \int P\{Z_{n+1} \leqq t \mid Z_2 = x, R_2 = r, Z_1 = 0\}\, dH(x, r). \end{aligned} \tag{2.3}$$

Since $\{Y_n\}$ $(Y_n = (Z_n, R_n))$ is a stationary Markov process and because of

Lemma 1

$$P\{Z_{n+1} \leqq t \mid Z_2 = x, R_2 = r, Z_1 = 0\} \\ = P\{Z_n \leqq t \mid Z_1 = x, R_1 = r\} \leqq P\{Z_n \leqq t \mid Z_1 = 0, R_1 = r\}. \tag{2.4}$$

Let H^* be the distribution function of R_1 and, therefore, of R_2. Then, using (2.4) in (2.3), we have

$$P\{Z_{n+1} \leqq t \mid Z_1 = 0\} \leqq \int P\{Z_n \leqq t \mid Z_1 = 0, R_1 = r\}\, dH(x, r)$$
$$= \int P\{Z_n \leqq t \mid Z_1 = 0, R_1 = r\}\, dH^*(r) = P\{Z_n \leqq t \mid Z_1 = 0\}.$$

Thus $P\{Z_n \leqq t \mid Z_1 = 0\}$ is a monotone sequence and therefore converges to a limit which we call $F(t)$. The above-mentioned properties of F are easily deduced.

THEOREM 1: *If* $Eg_2 > \max\,(ER_1, E\rho_1)$ *then* F *(defined in Lemma 2) is a probability distribution.*

PROOF: Because of Lemma 2 we need only show that, under the conditions stated here, $\{Z_n\}$ is bounded in probability, i.e., for all n

$$P\{Z_n \leqq t \mid Z_1 = 0\} \geqq 1 - \eta(t) \tag{2.5}$$

where $\eta(t) \to 0$ as $t \to \infty$. If we can prove that $\{W_n\}$ and $\{W_n^*\}$ are each bounded in probability then (2.5) will be established. Lindley [4] has shown that $\{W_n\}$ is bounded in probability so it remains only to consider $\{W_n^*\}$.

For $j = 1, 2, \cdots$ let

$$S_j = \sum_{i=1}^{j} (R_i - g_{i+1}) \tag{2.6}$$

and let $S_0 = 0$. By iterating (2.1) and using the fact that $W_1 = 0$ we have

$$W_{n+1} = \max_{0 \leq j \leq n} [S_n - S_j]. \tag{2.7}$$

Hence

$$R_n - g_{n+1} + W_n = R_n - g_{n+1} + \max_{0 \leq j \leq n-1} [S_{n-1} - S_j] = \max_{0 \leq j \leq n-1} [S_n - S_j]. \tag{2.8}$$

For $k \geqq 0$, let

$$B_k = \max_{0 \leq j \leq k} (-S_j). \tag{2.9}$$

Then (2.7), (2.8), and (2.9) yield

$$R_n - g_{n+1} + W_n - W_{n+1} = B_{n-1} - B_n. \tag{2.10}$$

Using (2.10) in (2.2) gives

$$W_{n+1}^* = \max\,[0, W_n^* + \rho_n - R_{n+1} + B_{n-1} - B_n]. \tag{2.11}$$

Put

$$(2.12)\qquad T_k = \sum_{i=1}^{k} (\rho_i - R_{i+1}) \quad \text{for} \quad k \geqq 1 \quad \text{and} \quad T_0 = 0.$$

Iterating (2.11) (use $W_1^* = 0$) we have

$$(2.13)\qquad \begin{aligned} W_{n+1}^* &= \max_{0 \leq k \leq n} [T_n - T_k + B_k - B_n] \\ &= \max_{0 \leq k \leq n} [T_n - T_k + \max_{0 \leq j \leq k} (-S_j) - B_n] \\ &= \max_{0 \leq j \leq k \leq n} [T_n - T_k - S_j - B_n]. \end{aligned}$$

Let $\epsilon \geqq 0$ (we shall specify ϵ later). Then

$$\begin{aligned} W_{n+1}^* &= \max_{0 \leq j \leq k \leq n} [T_n - T_k - (n-k)\epsilon + (n-k)\epsilon - S_j - B_n] \\ (2.14)\quad &\leqq \max_{0 \leq j \leq k \leq n} [T_n - T_k - (n-k)\epsilon] + \max_{0 \leq j \leq k \leq n} [(n-k)\epsilon - S_j - B_n] \\ &= \max_{0 \leq k \leq n} [T_n - T_k - (n-k)\epsilon] + \max_{0 \leq j \leq n} [(n-j)\epsilon - S_j - B_n]. \end{aligned}$$

Define $\xi_i = \rho_i - R_{i+1} - \epsilon$ and let $U_k = \sum_{i=1}^{k} \xi_i$. Thus U_k is the kth partial sum of independent and identically distributed random variables. Then

$$(2.15)\qquad \max_{0 \leq k \leq n} [T_n - T_k - (n-k)\epsilon] = \max_{0 \leq k \leq n} (U_n - U_k) = A_n \quad (\text{say})$$

has the same distribution as $\max_{0 \leq k \leq n} U_k$. If ϵ is such that

$$(2.16)\qquad E\rho_1 - ER_1 - \epsilon < 0$$

then, it is well known, $\max_{0 \leq k \leq n} U_k \to$ a finite random variable with probability one (w.p.1) which implies that $\{A_n\}$ is bounded in probability.

Observe that $B_n = \max_{0 \leq j \leq n} (-S_j) \geqq -S_n$. Hence

$$(2.17)\qquad \max_{0 \leq j \leq n} [(n-j)\epsilon - S_j - B_n] \leqq \max_{0 \leq j \leq n} [S_n - S_j + (n-j)\epsilon] = C_n \quad (\text{say}).$$

Let $V_j = \sum_{i=1}^{j} (R_i - g_{i+1} + \epsilon)$. Then V_j is the jth partial sum of independent and identically distributed random variables. Thus, as before, C_n has the same distribution as $\max_{0 \leq j \leq n} V_j$ and, if ϵ is such that

$$(2.18)\qquad ER_1 - Eg_2 + \epsilon < 0,$$

then $\{C_n\}$ is bounded in probability. Since $W_{n+1}^* < A_n + C_n$ we have only to verify that ϵ can be chosen to satisfy (2.16) and (2.18) in order to conclude that $\{W_n^*\}$ is bounded in probability.

If $E\rho_1 < ER_1$ then the choice $\epsilon = 0$ gives (2.16) and the condition of the Theorem guarantees (2.18). If $E\rho_1 \geqq ER_1$ take

$$\epsilon = [(E\rho_1 + Eg_2)/2] - ER_1$$

ϵ is clearly positive and (2.16) and (2.18) are satisfied because $Eg_2 > E\rho_1$. This concludes the proof of Theorem 1.

Theorem 2 which we now prove shows the necessity of the condition of Theorem 1 when first moments are assumed to exist.

THEOREM 2: (a) *If* $ER_1 \geqq Eg_2$ *then* $F(t) \equiv 0$.
(b) *If* $E\rho_1 \geqq Eg_2$ *then* $F(t) \equiv 0$.

PROOF: (a) is due to Lindley [4] who proved that, in this case, $W_n \to +\infty$ in probability. For (b) we might as well assume in addition that $ER_1 < Eg_2$ otherwise we can use (a).

If $ER_1 < Eg_2$ then

$$-S_n - B_n = -\max_{0 \leq j \leq n} [S_n - S_j] \tag{2.19}$$

is bounded in probability. From (2.13)

$$\begin{aligned} W^*_{n+1} &= \max_{0 \leq j \leq k \leq n} [T_n - T_k - S_j - B_n] \\ &= \max_{0 \leq j \leq k \leq n} [T_n + S_n - T_k - S_j] - S_n - B_n. \end{aligned} \tag{2.20}$$

Now, recalling that $R_1 > 0$ w.p.1,

$$\begin{aligned} \max_{0 \leq j \leq k \leq n} [T_n + S_n - T_k - S_j] &\geqq \max_{0 \leq k \leq n} [T_n + S_n - T_k - S_k] \\ &= \max_{0 \leq k \leq n} \left[\sum_{i=k+1}^{n} (\rho_i - g_{i+1}) + R_{k+1} - R_{n+1} \right. \\ &\geqq \max_{0 \leq k \leq n} \left[\sum_{i=k+1}^{n} (\rho_i - g_{i+1}) \right] - R_{n+1} \end{aligned} \tag{2.21}$$

$\{R_{n+1}\}$ is, of course, bounded in probability but, because $E\rho_1 - Eg_2 \geqq 0$,

$$\max_{0 \leq k \leq n} \left[\sum_{i=k+1}^{n} (\rho_i - g_{i+1}) \right] \to +\infty \quad \text{in probability.} \tag{2.22}$$

(2.22), (2.21) and (2.19) show that $W^*_{n+1} \to +\infty$ in probability which proves that $F(t) \equiv 0$. This proves Theorem 2.

It is interesting to note that if $ER_1 \geqq Eg_2$ and $E\rho_1 < ER_1$ then, although $W_n \to +\infty$ in probability, W^*_n has a legitimate limiting distribution. This is because we can show (as in Lemma 2) that $P\{W^*_n \leqq t_1 \mid Z_1 = 0\}$ has a limit and because of (2.13)

$$\begin{aligned} W^*_{n+1} &\leqq \max_{0 \leq k \leq n} [T_n - T_k] + \max_{0 \leq k \leq n} [B_k - B_n] \\ &= \max_{0 \leq k \leq n} [T_n - T_k] \end{aligned}$$

which is bounded in probability.

Just as in Kiefer and Wolfowitz [3] we can write down an integral equation for the limiting distribution of Y_n. Under the conditions of Theorem 1 this integral equation will have a unique probability distribution as a solution. The uniqueness argument in [3] is rather delicate but in this problem the difficulty is easily disposed of because of the ease in seeing (by means of (2.13) for example) that the limiting distribution must be independent of the starting point (W_1, W_1^*).

3. The s-Server case. The question of ergodicity in the case of s servers can be handled in essentially the same fashion as in Section 2 where we had 2 servers. We shall be brief in those places where the generalization of the ideas in Section 2 is transparent.

For $\sigma = 1, \cdots, s$ let R_n^σ be the service time in server σ of the nth person. Let $R_{n+1}^0 = \tau_{n+1} - \tau_n$ where τ_n is the time at which the nth person arrives to the first queue. Let $t_n^\sigma = R_n^\sigma - R_{n+1}^{\sigma-1}$, $\sigma = 1, \cdots, s$. Let W_n^σ be the waiting time in the σth queue of the nth person. It is easily verified that, for all $1 \leqq p \leqq s$,

$$W_{n+1}^p = \max\left[0, W_n^p + t_n^p + \sum_{\sigma=1}^{p-1} [t_n^\sigma + W_n^\sigma - W_{n+1}^\sigma]\right]. \tag{3.1}$$

Let $T_k^\sigma = \sum_{i=1}^k t_i^\sigma$ and let $D_k^p = \max^* [-T_{j_1}^1 - \cdots - T_{j_p}^p]$ where $\max^*$ is maximum over all $0 \leqq j_1 \leqq \cdots \leqq j_p \leqq k$. Let $H_n^p = \sum_{\sigma=1}^p [t_n^\sigma + W_n^\sigma - W_{n+1}^\sigma]$. To obtain a manageable expression for W_{n+1}^p we will show that $H_n^p = D_{n-1}^p - D_n^p$ for all p, n. Observe first that this is true when $p = 1$ and all n. Assume now that $H_n^p = D_{n-1}^p - D_n^p$ for all n. We will show that $H_n^{p+1} = D_{n-1}^{p+1} - D_n^{p+1}$ for all n.

From (3.1), the induction hypothesis, and iteration

$$\begin{aligned} W_{k+1}^{p+1} &= \max[0, W_k^{p+1} + t_k^{p+1} + H_k^p] = \max[0, W_k^{p+1} + t_k^{p+1} + D_{k-1}^p - D_k^p] \\ &= \max_{0 \leqq j \leqq k} [T_k^{p+1} - T_j^{p+1} + D_j^p - D_k^p]. \end{aligned} \tag{3.2}$$

Hence, using (3.2) for $k = n - 1$ and $k = n$

$$\begin{aligned} t_n^{p+1} + W_n^{p+1} - W_{n+1}^{p+1} &= t_n^{p+1} + \max_{0 \leqq j \leqq n-1} [T_{n-1}^{p+1} - T_j^{p+1} + D_j^p - D_{n-1}^p] \\ &\quad - \max_{0 \leqq j \leqq n} [T_n^{p+1} - T_j^{p+1} + D_j^p - D_n^p] \\ &= \max_{0 \leqq j \leqq n-1} [-T_j^{p+1} + D_j^p] - \max_{0 \leqq j \leqq n} [-T_j^{p+1} + D_j^p] \\ &\quad + D_n^p - D_{n-1}^p = D_{n-1}^{p+1} - D_n^{p+1} + D_n^p - D_{n-1}^p. \end{aligned}$$

Thus

$$H_n^{p+1} = t_n^{p+1} + W_n^{p+1} - W_{n+1}^{p+1} + H_n^p = D_{n-1}^{p+1} - D_n^{p+1}$$

which is what we wanted to show. Since H_n^p is what we say it is (3.2) is valid for all k and p (of course since there are only s servers we have no use for W_k^p where $p > s$).

Returning to (3.1) we remark that it is easy to verify just as in Section 2 that $Y_n = (W_n^1, \cdots, W_n^s, R_n^1, \cdots, R_n^{s-1})$ is the nth random variable in a stationary Markov process and that

$$P\{W_n^1 \leqq \alpha_1, \cdots, W_n^s \leqq \alpha_s \mid W_1^\sigma = 0, \sigma = 1, \cdots, s\} \to F(\alpha_1, \cdots, \alpha_s) \tag{3.3}$$

where F is an s-dimensional distribution function but not necessarily a probability distribution function.

Let $\mu_\sigma = ER_1^\sigma \qquad \sigma = 0, 1, \cdots, s$.

THEOREM 3: *If*

$$\max_{1 \leqq \sigma \leqq s} \mu_\sigma < \mu_0 \tag{3.4}$$

then F is a probability distribution.

PROOF: As in Theorem 1 we only have to show that each $\{W_n^\sigma\}$ is bounded in probability. It is easy to see by a trivial induction argument that we only have to verify that $\{W_n^s\}$ is bounded in probability. Actually the argument we give is legitimate when s is replaced by p for any $1 \leqq p \leqq s$. In any case we will only consider $\{W_n^s\}$.

To begin with observe that

$$-D_n^{s-1} \leqq T_n^1 + \cdots + T_n^{s-1}. \tag{3.5}$$

Hence from (3.2) with $k = n$, $p = s - 1$

$$\begin{aligned} W_{n+1}^s &\leqq \max_{0 \leqq j \leqq n} [T_n^1 + \cdots + T_n^s - T_j^s + D_j^{s-1}] \\ &= \max_{0 \leqq j_1 \leqq \cdots \leqq j_s \leqq n} [T_n^s - T_{j_s}^s + \cdots + T_n^1 - T_{j_1}^1] \end{aligned} \tag{3.6}$$

Let $s_0 = s$ and define s_i to be the largest $\sigma < s_{i-1}$ ($\sigma \geqq 0$) with the property that

$$\mu_\sigma - \mu_{s_{i-1}} > 0. \tag{3.7}$$

$\sigma = 0$ satisfies (3.7) because of (3.4) so that s_1 is well-defined. Let k be the first i such that $s_i = 0$. Then it is easy to check that

$$\mu_0 = \mu_{s_k} > \mu_{s_{k-1}} > \cdots > \mu_{s_0} = \mu_s \tag{3.8}$$

and that for $s_i < \sigma < s_{i-1}$

$$\mu_\sigma \leqq \mu_{s_{i-1}} < \mu_{s_i}. \tag{3.9}$$

Define, for $i = 1, \cdots, k$,

$$\begin{aligned} U_n^i &= \max_{0 \leqq j_1 \leqq \cdots \leqq j_s \leqq n} \left[\sum_{\sigma = s_i + 1}^{s_{i-1}} (T_n^\sigma - T_{j_\sigma}^\sigma)\right] \\ &= \max_{0 \leqq j_{s_i+1} \leqq \cdots \leqq j_{s_{i-1}} \leqq n} \left[\sum_{\sigma = s_i+1}^{s_{i-1}} (T_n^\sigma - T_{j_\sigma}^\sigma)\right]. \end{aligned} \tag{3.10}$$

Because of (3.6) we have

$$W_{n+1}^{*} \leq U_n^1 + \cdots + U_n^k \tag{3.11}$$

and, therefore, in order to show that $\{W_n^*\}$ is bounded in probability, we have only to verify that each $U_n^i(i = 1, \cdots, k)$ is bounded in probability.

The verification that each U_n^i is bounded can be summarized in the following lemma.

LEMMA: *For $m = 1, \cdots, M$ let $\{X_i^m, i = 1, \cdots\}$ be a sequence of independent and identically distributed random variables with*

$$EX_1^m = \lambda_m - \lambda_{m-1} \tag{3.12}$$

where

$$\lambda_0 > \lambda_M \geq \max_{0<s<M} \lambda_s > \min_{0<s<M} \lambda_s \geq 0. \tag{3.13}$$

(It is not assumed that $\{X_i^m\}$ and $\{X_i^{m'}\}$ are independent of one another). Let $S_k^m = \sum_{i=1}^k X_i^m$ and let $\psi_n = \max^* [\sum_{m=1}^M (S_n^m - S_{j_m}^m)]$ where $\max^*$ is maximum over all $j_1, \cdots, j_M$ with $0 \leq j_1 \leq \cdots \leq j_M \leq n$. Then ψ_n is bounded in probability.

It is easy to verify that U_n^i can be taken as ψ_n and that the conditions of the lemma are satisfied ((3.9) giving (3.13)) so that the proof of the lemma is the last step in proving Theorem 3.

PROOF OF LEMMA: Let $\gamma = \min' [\lambda_i - \lambda_j]$ where $\min'$ is minimum over all $0 \leq i, j \leq M$ with $\lambda_i - \lambda_j > 0$. Let $\delta = \gamma/M$. δ is, of course, strictly positive. For $2 \leq m \leq M$ define $\epsilon_m = \lambda_M - \lambda_{m-1} + (M - m + 1)\delta$ and let $\epsilon_{M+1} = 0$ and $\epsilon_1 = 0$. Then

$$\epsilon_2, \cdots, \epsilon_M \text{ are positive,} \tag{3.14}$$

$$\lambda_m - \lambda_{m-1} + \epsilon_{m+1} - \epsilon_m = -\delta, \qquad 2 \leq m \leq M \tag{3.15}$$

$$\lambda_1 - \lambda_0 + \epsilon_2 = \lambda_M - \lambda_0 + [(M-1)/M]\gamma < (-1/M)(\lambda_0 - \lambda_M) < 0. \tag{3.16}$$

Letting $j_{M+1} = n$ and taking note of the fact that $\epsilon_{M+1} = 0$ we have

$$\sum_{m=1}^{M} [(n - j_{m+1})\epsilon_{m+1} - (n - j_m)\epsilon_m] = 0 \tag{3.17}$$

where $0 \leq j_1 \leq j_2 \leq \cdots \leq j_M \leq j_{M+1} = \text{n}$. Now, using (3.17),

$$\psi_n = \max^* \left[\sum_{m=1}^{M} (S_n^m - S_{j_m}^m + (n - j_{m+1})\epsilon_{m+1} - (n - j_m)\epsilon_m) \right]$$

$$\leq \sum_{m=1}^{M} \max^* [S_n^m - S_{j_m}^m + (n - j_{m+1})\epsilon_{m+1} - (n - j_m)\epsilon_m] \tag{3.18}$$

$$\leq \sum_{m=1}^{M} \max_{0 \leq j_m \leq n} [S_n^m - S_{j_m}^m + (n - j_m)(\epsilon_{m+1} - \epsilon_m)].$$

Each of the terms in the summation on the right hand side of (3.18) is bounded

in probability since

$$\max_{0\le j_m\le n} [S_n^m - S_{j_m}^m + (n - j_m)(\epsilon_{m+1} - \epsilon_m)] = \max_{0\le k\le n} [\xi_n^m - \xi_k^m]$$

where $\xi_k^m = \sum_{i=1}^k (X_i^m + \epsilon_{m+1} - \epsilon_m)$ and $E(X_i^m + \epsilon_{m+1} - \epsilon_m) < 0$ due to to (3.15) and (3.16). This concludes the proof of the lemma and, therefore, the theorem.

THEOREM 4: *If* $\max_{1\le\sigma\le s}\mu_\sigma \geqq \mu_0$ *F is identically* 0.

PROOF: Letting p be the first $\sigma > 1$ with $\mu_\sigma \geqq \mu_0$ we need only show that $W_n^p \to +\infty$ in probability. Using (3.2)

$$\begin{aligned} W_{n+1}^p &= \max_{0\le j\le n} [T_n^p - T_j^p + D_j^{p-1} - D_n^{p-1}] \\ &= \max_{0\le j_1\le\cdots j_p\le n} [T_n^p - T_{j_p}^p - T_{j_{p-1}}^{p-1} - \cdots - T_{j_1}^1 - D_n^{p-1}] \\ &= \max_{0\le j_1\le\cdots\le j_p\le n} [T_n^p + \cdots + T_n^1 - T_{j_p}^p - \cdots - T_{j_1}^1] \\ &\quad - \max_{0\le j_1\le\cdots\le j_p\le n} [T_n^1 + \cdots + T_n^{p-1} - T_{j_1}^1 - \cdots - T_{j_{p-1}}^{p-1}]. \end{aligned} \tag{3.19}$$

The last term on the right hand side of (3.19) is bounded in probability because $\max_{1\le\sigma\le p-1}\mu_\sigma < \mu_0$. Looking at the first term on the right hand side of (3.19) we have

$$\begin{aligned} &\max_{0\le j_1\le\cdots\le j_p\le n} [T_n^p + \cdots + T_n^1 - T_{j_p}^p - \cdots - T_{j_1}^1] \\ &\qquad = \max_{0\le k\le n}\left[\sum_{j=k+1}^n \sum_{\sigma=1}^p (R_j^\sigma - R_{j+1}^{\sigma-1})\right] \\ &\qquad = \max_{0\le k\le n}\left[\sum_{j=k+1}^n (R_j^p - R_j^0) + \sum_{\sigma=1}^p (R_{k+1}^{\sigma-1} - R_{n+1}^{\sigma-1})\right] \\ &\qquad \geqq \max_{0\le k\le n}\left[\sum_{j=k+1}^n (R_j^p - R_j^0)\right] - \sum_{\sigma=1}^p R_{n+1}^{\sigma-1}. \end{aligned}$$

The last term written is bounded in probability while the preceding term goes to $+\infty$ in probability because $\mu_p \geqq \mu_0$. It is then quite clear that W_{n+1}^p must go to $+\infty$ in probability.

REFERENCES

[1] HIROTUGU AKAIKE, "On ergodic property of a tandem type queueing process," *Ann. Inst. Stat. Math.* (*Tokyo*), Vol. 9 (1957), pp. 13–22.

[2] PAUL J. BURKE, "The output of a queueing system," *Operations Research*, Vol. 4 (1956), pp. 699–704.

[3] J. KIEFER AND J. WOLFOWITZ, "On the theory of queues with many servers," *Trans. Am. Math. Soc.*, Vol. 78, (1955), pp. 1–18.

[4] D. V. LINDLEY, "The theory of queues with a single server," *Proc. Cambridge Philos. Soc.*, Vol. 48 (1952), pp. 277–289.

[5] EDGAR REICH, "Waiting times when queues are in tandem," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 768–772.

QUEUES FOR A FIXED-CYCLE TRAFFIC LIGHT

By G. F. Newell[1]

Brown University

1. Summary. In their book *Studies in the Economics of Transportation*, Beckmann, McGuire and Winsten (BMW) ([2], pp. 11–13, 40–42) proposed a simple queuing model for traffic flow through a fixed-cycle traffic light. Although they derived a relation between the average delay per car and the average length of the queue at the beginning of a red phase of the light, they only indicated some possible numerical schemes for evaluating the latter. Here we shall derive analytic expressions for the average queue length and consequently also the average delay under equilibrium conditions for the BMW model.

2. Introduction. Several papers have been written on the subject of queuing at a fixed-cycle traffic light. Wardrop [7] and Webster [8] describe very extensive studies based upon experimental observation, computer simulation and semi-empirical theory with the theory based upon the assumption that the arrivals of cars at the light form a Poisson process. Uematu [6] investigated the queues for a model quite similar to that of BMW but was mainly concerned with the question of how long it takes an empty queue to reach some preassigned length for the first time. The present author also made a previous study [5] of delays but only considered arrival rates which were not too close to the critical value and used a more elaborate model than that considered here.

In the model proposed by BMW, it is assumed that events such as the arrival or departure of a car at the traffic light may occur only on a set of discrete and equally spaced time points. The traffic light pattern is periodic in time with each cycle represented by a sequence of r consecutive time points designated as red points followed by a sequence of g points designated as green. At either a red or green point there is a probability α that one new car will arrive and a probability $1 - \alpha$ that no new cars arrive, these probabilities being independent of the number of arrivals at any other time points. No cars are allowed to leave the light at red points but one car leaves at any green point provided that either a new car also arrives at that time or the queue just prior to this time point is non-empty.

From these rules it follows that the lengths of the queue immediately before time points define a non-stationary Markov chain in which at any red point there is a probability α that the queue increases by one car and a probability $1 - \alpha$ that it remains unchanged, whereas at any green point there is a probability α that a non-empty queue remains unchanged and a probability $1 - \alpha$ that it decreases by one. The lengths of the queue before corresponding time points of successive cycles of the light, however, form a stationary Markov

Received June 18, 1959; revised March 8, 1960.

[1] Alfred P. Sloan Research Fellow 1956–1959.

chain. If we let q_x denote the length of queue just before the first red point of the xth cycle, the q_x satisfy the recursion relation,

$$q_{x+1} = \max\{q_x + u_x - g, 0\}, \tag{2.1}$$

in which u_x represents the total number of arrivals during the xth cycle. The u_x are independent random variables having a binomial distribution

$$\Pr\{u_x = m\} = \binom{r+g}{m}(1-\alpha)^{r+g-m}\alpha^m . \tag{2.2}$$

Our problem here is to find the equilibrium distribution for q_x. Once this has been found and $E(q_x)$ evaluated, the average waiting time per car measured in units of the time interval between consecutive time points can be evaluated from the formula derived by BMW,

$$w = r(1-\alpha)^{-1}(g+r)^{-1}[E(q_x)/\alpha + (r+1)/2]. \tag{2.3}$$

Relation (2.1) is equivalent to the recursion formula for a queue with bulk service of g customers at a time. It has been studied previously by Bailey [1] and Downton [3], [4] when arrivals have a Poisson distribution and service time a χ^2 distribution (a special case of which is service at constant time intervals). Some of the analysis here for a binomial distribution of arrivals, particularly Section 4, closely parallels the analysis described by Bailey.

3. Low Rates of Arrival. One method of determining the equilibrium distribution of q_x is to take any initial distribution, for example $q_1 = 0$ with probability one, and evaluate the distribution for q_2, q_3, etc. from (2.1) and (2.2). If the average rate of arrivals per cycle is less than the maximum rate of departure, i.e.,

$$\alpha(r+g) < g \tag{3.1}$$

then this sequence of distributions will always converge to the equilibrium distribution.

If, in addition to (3.1), the difference between these rates is larger than the dispersion of u_x, i.e.,

$$g - \alpha(r+g) > [\alpha(1-\alpha)(r+g)]^{\frac{1}{2}}, \tag{3.2}$$

then $\Pr\{q_2 > 0\}$ will be small compared with $\Pr\{q_2 = 0\}$, $\Pr\{q_2 > 0 \text{ and } q_3 > 0\}$ will be relatively much smaller yet, and the sequence of distributions for q_2, q_3, etc. will converge rapidly to the equilibrium distribution, the more rapidly the larger the difference in the two sides of (3.2).

If we take $\Pr\{q_1 = 0\} = 1$, the next approximation to the equilibrium distribution is

$$\begin{aligned} \Pr\{q_2 = j\} &= \binom{r+g}{g+j}(1-\alpha)^{r-j}\alpha^{g+j}, \qquad j > 0, \\ \Pr\{q_2 = 0\} &= 1 - \sum_{j=1}^{r} \Pr\{q_2 = j\}. \end{aligned} \tag{3.3}$$

The evaluation of the distributions for q_3, q_4, etc. is straightforward but becomes quite tedious.

Since, in most practical applications, r and g are in the range of 10 to 20, we expect that estimations of $E(q_s)$ in the limit r and $g \to \infty$ with r/g fixed will be of some value. In this limit, (3.3) can be used to approximate $E(q_s)$ whenever

$$\mu \equiv [g - \alpha(r + g)][rg/(r + g)]^{-\frac{1}{2}} > 1, \tag{3.4}$$

a condition which excludes only a range of α in which the difference between α and the critical value, $g/(r + g)$, is of order $r^{-\frac{1}{2}}$. For r sufficiently large, this excluded range can be made arbitrarily small but if $r = g = 10$, for example, it is from $\alpha \sim 0.38$ to 0.5 and for $r = g = 20$ from $\alpha \sim 0.42$ to 0.5.

From (3.3) we obtain for $0 < j \ll r$

$$\begin{aligned} &\Pr\{q_2 = j\} \\ &= \left\{\frac{(r + g)!(1 - \alpha)^r\alpha^g}{g!r!}\right\}\left\{\frac{r\alpha}{g(1 - \alpha)}\right\}^j \exp\left\{\frac{-j^2(r + g)}{2rg} + O\left(\frac{j^3}{r^2}\right)\right\} \end{aligned} \tag{3.5}$$

and

$$E(q_2) = \frac{(r + g + 1)!(1 - \alpha)^{r+1}\alpha^{g+1}}{g!r!\mu^2}[1 + O(\mu^{-2})]. \tag{3.6}$$

If we disregard the smaller values of α and assume that $\mu \ll r^{1/6}$, then (3.6) can be simplified further by using Stirling's formula and expansions of $\log \alpha$ in powers of μ to give

$$E(q_2) = \left[\frac{gr}{2\pi(r + g)}\right]^{\frac{1}{2}} \frac{\exp(-\mu^2/2)}{\mu^2}\left[1 + O\left(\frac{\mu^3}{r^{\frac{1}{2}}}\right) + O(\mu^{-2})\right]. \tag{3.7}$$

For $\mu^2 \gg 1$, we can also estimate that $E(q_3)$ will differ from $E(q_2)$ only by an additional term that is smaller than $E(q_2)$ by a factor proportional to $\exp(-\mu^2/2)$.

Whereas in practical applications, the error terms in (3.6) or (3.7) may be quite significant, these equations at least give an accurate description of what happens for sufficiently large r and μ and a qualitative description even for moderately large r. In the range $\mu > 1$, $E(q_2)$ is a monotone increasing function of α and is of order $r^{\frac{1}{2}}$ for $\mu = O(1)$. For $r \gg 1$, μ is a rapidly varying function of α and as α decreases $E(q_2)$ also decreases very rapidly. Even for the largest α at which we may apply these formulas, however, where $E(q_s)$ is of order $r^{\frac{1}{2}}$, the effect of the queue on w is small because in (2.3) $E(q_s)$ must be added to another term that is of order r. For $x = g = 10$, $E(q_s)$ causes only about a 20% increase in w even when $\mu = 1$.

To investigate what happens for $\mu < 1$, we consider below a different method of evaluating $E(q_s)$

4. Use of Generating Functions. Let

$$G_x(z) = \sum_{j=0}^{\infty} z^j \Pr\{q_x = j\} \tag{4.1}$$

denote the probability generating function (p.g.f.) for q_x. From (2.2) u_x has the p.g.f. $(1 - \alpha + \alpha z)^{r+g}$ and, since u_x and q_x are independent, $u_x + q_x - g$ has the p.g.f. $(1 - \alpha + \alpha z)^{r+g} z^{-g} G_x(z)$. If we subtract from this the probabilities for negative values of $u_x + q_x - g$ and reassign them to the event $q_{x+1} = 0$, we obtain from (2.1) the p.g.f. for q_{x+1},

$$G_{x+1}(z) = z^{-g}\left[(1 - \alpha + \alpha z)^{r+g} G_x(z) - \sum_{k=0}^{g-1} a_k z^k\right] + \sum_{k=0}^{g-1} a_k, \tag{4.2}$$

in which the a_k are the Taylor expansion coefficients of $(1 - \alpha + \alpha z)^{r+g} G_x(z)$.

If there is an equilibrium distribution for the queue length with $G_{x+1}(z) = G_x(z) \equiv G(z)$ then (4.2) gives

$$G(z) = Q^{-1}(z)\left[z^g \sum_{k=0}^{g-1} a_k - \sum_{k=0}^{g-1} a_k z^k\right] \tag{4.3}$$

with

$$Q(z) = z^g - (1 - \alpha + \alpha z)^{r+g}. \tag{4.4}$$

We do not know the a_k unless we know $G(z)$, but (4.3) and (4.4) at least describe the form of $G(z)$, a polynomial of degree g divided by another polynomial of degree $r + g$. We also know that, if $G(z)$ is a p.g.f., it must be analytic in the unit circle $|z| \leqq 1$ of the complex plane, and in particular at any points in this circle where $Q(z)$ has a zero.

Since $Q(z)$ is analytic, the number of zeros of $Q(z)$ inside or on the circle $|z| = 1$ is equal to g plus the number of cycles through which the complex phase of $z^{-g}Q(z)$ changes when z traverses a path just outside the unit circle, or equivalently g plus the number of times the image of this path under the transformation $z^{-g}Q(z)$ encircles the origin. Since for $|z| = 1$ and $0 < \alpha < 1$

$$|z^{-g}Q(z) - 1| = |1 - \alpha + \alpha z|^{r+g} \leqq 1,$$

with the last equality sign valid only at $z = 1$, the image of the unit circle itself passes through the origin once as z passes through $z = 1$ but otherwise lies to the right of the origin. Whether or not $z^{-g}Q(z)$ encircles the origin as z traverses a path just outside the unit circle is, therefore determined by what happens to $z^{-g}Q(z)$ for z in the neighborhood of $z = 1$. By expanding $z^{-g}Q(z)$ in a Taylor series about $z = 1$, one can easily show that as z passes to the right of $z = 1$, $z^{-g}Q(z)$ passes to the right of the origin if $\alpha(r + g) < g$ and so fails to encircle the origin but passes to the left of the origin thereby encircling it once if $\alpha(r + g) \geqq g$. We conclude from this that $Q(z)$ has g zeros inside or on the unit circle if $\alpha(r + g) < g$ but $g + 1$ zeros if $\alpha(r + g) \geqq g$. Since $\alpha(r + g) < g$

is also the condition for existence of an equilibrium distribution of q_x, only this case is of interest here.

If $G(z)$ is to be analytic for $|z| \leqq 1$, each of the g factors $(z - z_l)$ of $Q(z)$ with $|z_l| \leqq 1$ must cancel a corresponding factor of the gth degree polynomial in the numerator of $G(z)$ and $G(z)$ will reduce to the form

$$G(z) = A \prod_{l=1}^{r} (z - z_l)^{-1},$$

in which z_l, $l = 1, 2, \cdots, r$ are the r zeros of $Q(z)$ with $|z_l| > 1$. Since, in addition, any p.g.f. must satisfy the condition $G(1) = 1$, we finally obtain

$$G(z) = \prod_{l=1}^{r} (1 - z_l)(z - z_l)^{-1} \tag{4.5}$$

and

$$E(q_x) = dG(z)/dz\,|_{z=1} = \sum_{l=1}^{r} (z_l - 1)^{-1}. \tag{4.6}$$

The study of the q_x distribution is thus reduced to a study of the roots z_l of $Q(z)$ with $|z_l| > 1$.

It is not generally possible to obtain explicit expressions for the roots z_l, but they must all lie on a curve of the complex plane defined by the equation

$$|z| = |1 - \alpha + \alpha z|^{(r+g)/g}. \tag{4.7}$$

For any specified direction of z in the complex plane, one can sketch the graphs of the two sides of (4.7) as a function of $|z|$ and show that for $\alpha(r + g) < g$, the two graphs always intersect twice, once for $|z| \leqq 1$ and once for $|z| > 1$. The curve of (4.7), therefore, consists of two closed paths C' and C such as shown in Fig. 1, one lying inside the unit circle and the other outside.

The roots z_l must also satisfy the equation

$$z_l^{g/r} = \gamma_l(1 - \alpha + \alpha z_l)^{(r+g)/r}, \tag{4.8}$$

with $\gamma_l^r = 1$, and one can show that there is one and only one root of (4.8) on the curve C of Fig. 1 corresponding to each of the r distinct values of γ_l with $\gamma_l^r = 1$. By suitable numbering of the roots z_l we can choose γ_l so that

$$\gamma_l = \exp[2\pi i(l - 1)/r]. \tag{4.9}$$

We can also interpret (4.8) as a one-to-one mapping of the r roots on C into the r values of γ_l equally spaced around the unit circle. If we let r and $g \to \infty$ keeping r/g and α fixed, the curve C also stays fixed but the values of γ_l become densely and uniformly distributed on the unit circle. At the same time, the roots z_l become dense on C.

We know already from Section 3 that for the above limiting process $E(q_x) \to 0$. This can also be derived from (4.6) by observing that for $r \to \infty$ the sum in (4.6) becomes the Riemann sum for an integral which we may interpret as

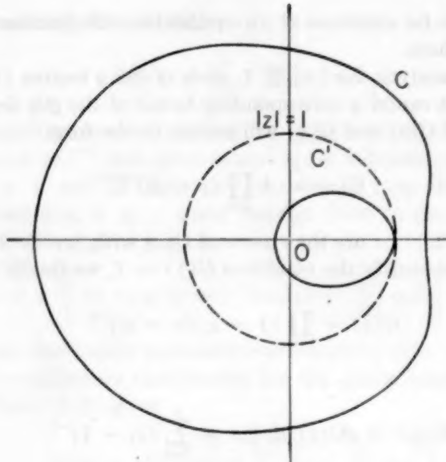


FIG. 1

either an integral with respect to the continuous real variable l or with respect to the complex variable γ around the unit circle. If we choose the last form we find

$$(4.10) \qquad r^{-1}E(q_x) \to 1/2\pi i \int \{\gamma[z(\gamma) - 1]\}^{-1}\, d\gamma.$$

The function $z(\gamma)$ defined by (4.8) for $|\gamma| = 1$ is also defined for $|\gamma| > 1$. For $|\gamma| \geqq 1$, $z(\gamma)$ is analytic, $|z(\gamma)| > 1$, and is of order γ for $\gamma \to \infty$. The contour integral in (4.10) therefore vanishes by virtue of Cauchy's theorem. In addition, the difference between the Riemann sum of an analytic function and the integral over any closed path is asymptotically smaller than any finite power of the spacing between points. $E(q_x)$ is, therefore, smaller than any finite power of r^{-1} for $r \to \infty$.

If we define z_l for non-integer real l through (4.8) and (4.9), it follows that

$$\int_{\frac{1}{2}}^{r+\frac{1}{2}} (z_l - 1)^{-1}\, dl = 0.$$

By dividing this integral into r parts and subtracting it from (4.6), we can also write $E(q_x)$ as the difference between a Riemann sum and its limiting integral, i.e.

$$(4.11) \qquad E(q_x) = \sum_{l=1}^{r} \left\{ (z_l - 1)^{-1} - \int_{l-\frac{1}{2}}^{l+\frac{1}{2}} (z_l - 1)^{-1}\, dl \right\}.$$

5. Nearly Critical Arrival Rate. Equations (4.6) and (4.11) are particularly

well suited to the evaluation of $E(q_z)$ when $\alpha \to g/(r+g)$ because in this case we find that $z_1 \to 1$ and the one term of (4.6) for $l = 1$ becomes infinite. If, however, we let $r \to \infty$ and $\alpha \to g/(r+g)$ simultaneously then some of the neighboring roots to z_1, for example z_2 and z_r also approach 1.

Since z_l is defined by (4.8) and (4.9) also for negative values of l and is periodic in l with period r, we may consider l in the range $-r/2 < l \leqq r/2$, for example, so that the roots nearest to z_1 are those with small $|l|$, $l = \cdots -1, 0, 2, \cdots$. To locate these roots we take the logarithm of both sides of (4.8) and expand in powers of $(z_l - 1)$ and μ to obtain

$$-4\pi i(l-1)(r+g)r^{-1}g^{-1} - 2(z_l-1)(r+g)^{\frac{1}{2}}(rg)^{-\frac{1}{2}}\mu + (z_l-1)^2 + O[(z_l-1)^2\mu r^{-\frac{1}{2}}, \quad (z_l-1)^3] = 0.$$

The roots of this approximately quadratic equation with $|z_l| > 1$ are

$$z_l - 1 = (r+g)^{\frac{1}{2}}(rg)^{-\frac{1}{2}}\{\mu + [\mu^2 + 4\pi i(l-1)]^{\frac{1}{2}}\} + O[(z_l-1)^2] \tag{5.1}$$

and in particular

$$z_1 - 1 = 2(r+g)^{\frac{1}{2}}(rg)^{-\frac{1}{2}}\mu + O[(r+g)r^{-1}g^{-1}\mu^2]. \tag{5.2}$$

We conclude immediately from this that, if r and g are finite,

$$\begin{aligned} E(q_z) &= (z_1-1)^{-1} + O(1), && \text{for } \mu \to 0, \\ &= rg\{2(r+g)[g - \alpha(r+g)]\}^{-1} + O(1), && \text{for } \alpha \to g/(r+g), \end{aligned} \tag{5.3}$$

in which $O(1)$ here means order relative to μ as $\mu \to 0$ but not relative to r and g.

Suppose we now let $r \to \infty$ with μ fixed, particularly with $\mu \lesssim 1$ since this is the only case that could not be handled satisfactorily in Section 3. Except when $l \ll r$ and $z_l \sim 1$, the difference between $(z_l - 1)^{-1}$ and its integral between $l - \frac{1}{2}$ and $l + \frac{1}{2}$ is of the order of magnitude of the second derivative of $(z_l - 1)^{-1}$ with respect to l, which in turn is of order r^{-2} according to (4.8) and (4.9). The sum of all such terms in (4.11) is at most of order r^{-1} and so any significant contribution to (4.11) can come only for the small values of $|l|$ where (5.1) is applicable. From (4.11) and (5.1), we obtain

$$E(q_z) = \left(\frac{rg}{r+g}\right)^{\frac{1}{2}} \sum_{l=-\infty}^{+\infty} \left\{ (\mu + [\mu^2 + 4\pi i(l-1)]^{\frac{1}{2}})^{-1} - \int_{l-\frac{1}{2}}^{l+\frac{1}{2}} (\mu + [\mu^2 + 4\pi i(l-1)]^{\frac{1}{2}})^{-1}\,dl \right\} + O(1) \tag{5.4}$$

with the dominant error team of $O(1)$ relative to r coming from the error term of (5.1), particularly for $l = 1$ and to a lesser extent from the other l with $|l| \ll r$.

The terms in the sum (5.4) are $O(l^{-\frac{3}{2}})$ for $|l| \gg \mu$, so the series converges rapidly enough to be of practical use even for $\mu \sim (4\pi)^{\frac{1}{2}} \sim 3$. For small μ, the main contribution, however, comes from $l = 1$ where the first term in the bracket of (5.4) is $(2\mu)^{-1}$ while all other contributions to the series are at most of order

1 even for $\mu \to 0$. Generally we obtain for μ of order 1 or less

$$(5.5) \qquad E(q_x) = [rg/(r+g)]^{\frac{1}{2}}[(2\mu)^{-1} + O(1)]$$

and for $\mu \lesssim 1$ we can expand (5.4) in powers of μ to obtain

$$(5.6) \qquad E(q_x) = \left(\frac{rg}{r+g}\right)^{\frac{1}{2}}\left[\frac{1}{2\mu} - A + \frac{\mu}{4} + O(\mu^2)\right]$$

with

$$(5.7) \qquad A = (2\pi)^{-\frac{1}{2}} \lim_{R\to\infty} [2(R+\tfrac{1}{2})^{\frac{1}{2}} - \sum_{l=1}^{R} l^{-\frac{1}{2}}] \sim 0.582.$$

One can estimate that $O(\mu^2)$ is roughly $-\mu^2/20$ and in succeeding terms the important parameter is $\mu/(4\pi)^{\frac{1}{2}}$, so that (5.6) will be correct to within about 30% even for $\mu = 1$. The error in (3.7) for $\mu = 1$ should be of comparable size and if one compares (5.6) with (3.7) one finds that they agree to within a factor of about $\frac{3}{2}$ for $r \to \infty$ and $\mu = 1$.

Since for $r \to \infty$, the effect of the queue on w will not be significant unless $E(q_x)$ is of order r, this will not occur unless μ is $O(r^{-\frac{1}{2}})$ and $g - \alpha(r+g) = O(1)$. If, in fact, $g - \alpha(r+g) = 1$ the queue causes w to increase by a factor of 2.

We note finally that for certain values of $(r+g)/r$, namely 2, $\frac{3}{2}$, 3, $\frac{4}{3}$ and 4, one can obtain exact expressions for the roots z_l by virtue of the fact that (4.8) gives a set of quadratic, cubic or quartic equations. One can, therefore, also obtain exact explicit formulas for $E(q_x)$ and w. If, for example, $r = g$, then

$$(5.8) \qquad z_l - 1 = \tfrac{1}{2}\gamma_l^{-1}\alpha^{-2}\{1 - 2\alpha\gamma_l + [1 - 4\alpha(1-\alpha)\gamma_l]^{\frac{1}{2}}\}$$

in which the square root must be chosen in the right half of the complex plane to give $|z_l| > 1$. In particular

$$(5.9) \qquad z_1 - 1 = \alpha^{-2}(1 - 2\alpha).$$

6. Comparison with Webster's Formula. The only formula with which we can compare the above results is Webster's semi-empirical formula [8] for delays which is based upon the assumption that the arrivals form a Poisson distribution rather than a binomial distribution as assumed here. Webster's formula consists of three terms; the first is essentially the same as (2.3) with $E(q_x) = 0$ and represents the delay for regularly spaced arrivals; the second term is the delay that results from a queue when arrivals have a Poisson distribution but the service time is a constant equal to $(r+g)/g$ time intervals; and the third term is an empirical correction obtained by fitting curves to values calculated by computer simulation.

Since a Poisson distribution allows arbitrary small time intervals between arrivals, fluctuations may cause more cars to arrive in some green period than can leave. Because of this one finds for a Poisson distribution of arrivals that even when $r \to 0$ (no traffic light) one still has a queue and furthermore the average length of the queue becomes infinite as the arrival rate approaches the critical value. As pointed out by BMW, the binomial distribution has the ad-

vantage of forcing a minimum spacing between cars and so we avoid this unfortunate limiting behavior, even though this is accomplished in a somewhat artificial way wherein the spacings are confined to be integer multiples of the minimum spacing.

By using methods very similar to those described in Sections 1 to 5, it is possible also to compute the queue lengths and delays when the arrivals have a Poisson distribution, provided we assume that (2.1) still holds. We need only replace the p.g.f. for the binomial distribution of u_s by the corresponding expression for the Poisson distribution. By doing this one finds as the analogue of (5.5)

$$E(q_z) = \tfrac{1}{2}g[g - \alpha(r + g)]^{-1} + O(r^{\frac{1}{2}}), \tag{6.1}$$

the leading term of which is $(g + r)/r$ times as large as in (5.5). The average waiting time for nearly critical arrival rate is then given by

$$w = r^2/\{2(1 - \alpha)(g + r)\} + (g + r)/\{2[g - \alpha(r + g)]\} + O(r^{\frac{1}{2}}). \tag{6.2}$$

Furthermore, in (6.1) and (6.2), the $O(r^{\frac{1}{2}})$ are asymptotically proportional to $r^{\frac{1}{2}}$.

The first term of (6.2) is essentially the same as the first term of Webster's formula and the coefficient of $[g - \alpha(r + g)]^{-1}$ in the second term has the same value as in Webster's formula for $\alpha \to g/(r + g)$. The third term of Webster's formula, however, is not asymptotically proportional to $r^{\frac{1}{2}}$, nor does it indicate in any obvious way the importance of the magnitude of $g - \alpha(r + g)$ as compared with $[rg/(r + g)]^{\frac{1}{2}}$.

7. Acknowledgements. Most of this work was while the author was on sabbatical leave from Brown University visiting the Stockholm Högskola. He would like to express his appreciation to Professor Malmquist, Docent Dalenius and the staff of the Institute for Statistics for their hospitality. A revision of the manuscript was financed by a grant from General Motors Corporation.

REFERENCES

[1] NORMAN T. J. BAILEY, "On queueing processes with bulk service," *J. Roy. Stat. Soc., Ser. B*, Vol. 16 (1954), pp. 80–87.

[2] MARTIN BECKMANN, C. B. MCGUIRE AND CHRISTOPHER V. WINSTEN, *Studies in the Economics of Transportation*, Yale University Press, New Haven, Conn., 1956.

[3] F. DOWNTON, "Waiting time in bulk service queues," *J. Roy. Stat. Soc., Ser. B*, Vol. 17 (1955), pp. 256–261.

[4] F. DOWNTON, "On limiting distributions arising in bulk service queues," *J. Roy. Stat. Soc., Ser. B*, Vol. 18 (1956), pp. 265–274.

[5] G. F. NEWELL, "Statistical analysis of the flow of highway traffic through a signalized intersection," *Q. Appl. Math.*, Vol. 13 (1956), pp. 353–369.

[6] TOSIO UEMATU, "On the traffic control at an intersection controlled by the repeated fixed-cycle traffic light," *Ann. Inst. Stat. Math. Tokyo*, Vol. 9 (1958), pp. 87–107.

[7] JOHN GLEN WARDROP, "Some theoretical aspects of road traffic research," *Proc. Instn. Civ. Engrs. II*, Vol. 1 (1952), pp. 325–362.

[8] F. V. WEBSTER, "Traffic signal settings," Road Research Technical Paper No. 39, Her Majesty's Stationery Office, London, 1958.

PROBABILITY CONTENT OF REGIONS UNDER SPHERICAL NORMAL DISTRIBUTIONS, I[1]

By Harold Ruben

Columbia University

1. Introduction. The primary purpose of this series of papers is to attempt to lay the groundwork for a relatively well-rounded theory of the spherical normal distribution. Many distributional problems in mathematical statistics may be regarded as particular instances of one general problem, the determination of the probability content of geometrically well-defined regions in Euclidean N-space when the underlying distribution is centered spherical normal and has unit variance in any direction. Specifically then, we require for a definite region R

$$P(R) = (2\pi)^{-\frac{1}{2}N} \int_{\mathbf{x} \epsilon R} e^{-\frac{1}{2}\mathbf{x}'\mathbf{x}}\, d\mathbf{x}, \tag{1.1}$$

in which $\mathbf{x}' = (x_1, \cdots, x_N)$. The class of problems represented by (1.1) is a very broad one and the literature on it is correspondingly quite enormous and well-diffused. In fact, all the distributional problems which occur in the theory of sampling from multivariate normal populations may in principle be brought under our general heading. Thus, let $\mathbf{y}_i$, $i = 1, 2, \cdots, n$, denote n mutually independent k-dimensional vectors each of which is governed by the elementary probability density

$$p(\mathbf{y}) = (2\pi)^{-\frac{1}{2}k} |\mathbf{V}|^{-\frac{1}{2}} e^{-\frac{1}{2}\mathbf{y}'\mathbf{V}^{-1}\mathbf{y}}. \tag{1.2}$$

The joint probability density function for the n vectors is $\prod_1^n p(\mathbf{y}_i)$ and integrals of the form

$$\begin{aligned}(2\pi)^{-\frac{1}{2}nk} |\mathbf{V}|^{-\frac{1}{2}n} \int_{\mathbf{z} \epsilon T} \exp\left(-\tfrac{1}{2}\sum_1^n \mathbf{y}_i'\mathbf{V}^{-1}\mathbf{y}_i\right) \prod_1^n d\mathbf{y}_i \\ = (2\pi)^{-\frac{1}{2}N} |\mathbf{W}|^{-\frac{1}{2}} \int_{\mathbf{z} \epsilon T} \exp\left(-\tfrac{1}{2}\mathbf{z}'\mathbf{W}^{-1}\mathbf{z}\right) d\mathbf{z},\end{aligned} \tag{1.3}$$

where $\mathbf{z}$ is a partitioned vector, $\mathbf{W}$ is a partitioned matrix,

$$\mathbf{z} = \begin{bmatrix} \mathbf{y}_1 \\ \cdots \\ \mathbf{y}_2 \\ \cdots \\ \vdots \\ \cdots \\ \mathbf{y}_n \end{bmatrix}, \qquad \mathbf{W} = \begin{bmatrix} \mathbf{V} & \vdots & \mathbf{0} & \cdots & \vdots & \mathbf{0} \\ \cdots & & \cdots & & & \cdots \\ \mathbf{0} & \vdots & \mathbf{V} & \cdots & \vdots & \mathbf{0} \\ \cdots & & \cdots & & & \cdots \\ \mathbf{0} & \vdots & \mathbf{0} & \cdots & \vdots & \mathbf{V} \end{bmatrix}, \tag{1.4}$$

Received February 13, 1959; revised March 5, 1960.

[1] This research was sponsored in part by the Office of Naval Research under Contract Number Nonr-266 (33), Project Number NR 042-034. Reproduction in whole or part is permitted for any purpose of the United States Government.

$N = nk$ and T a specified region in Euclidean N-space, may be thrown in the form (1.1) by a linear orthogonal transformation chosen so as to orient the new axes along the axes of the ellipsoids of constant density of the distribution of $\mathbf{z}$, followed by a simple scaling transformation to convert the ellipsoids into spheres.

We recall a second and frequently more convenient method of reducing $\mathbf{z}'\mathbf{W}^{-1}\mathbf{z}$ to a sum of squares. By means of triangular resolution, $\mathbf{W}$ may be factored [1] in the form

$$\mathbf{W} = \mathbf{M}\mathbf{M}', \tag{1.5}$$

where the $N \times N$ matrix $\mathbf{M}$ is defined by

$$\mathbf{M} = \left[\begin{array}{c:cccc} \mathbf{L} & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \hdashline \mathbf{0} & \mathbf{L} & \cdots & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \cdots & \mathbf{L} \end{array}\right], \tag{1.6}$$

and $\mathbf{V} = \mathbf{L}\mathbf{L}'$, $\mathbf{L}$ denoting a $k \times k$ lower triangular matrix. On setting

$$\mathbf{z} = \mathbf{M}\mathbf{x}, \tag{1.7}$$

$$(2\pi)^{-\frac{1}{2}N} \mid \mathbf{W} \mid^{-\frac{1}{2}} \int_{\mathbf{z}\epsilon T} \exp(-\tfrac{1}{2}\mathbf{z}'\mathbf{W}^{-1}\mathbf{z})\, d\mathbf{z} = (2\pi)^{-\frac{1}{2}N} \int_{\mathbf{x}\epsilon R} \exp(-\tfrac{1}{2}\mathbf{x}'\mathbf{x})\, d\mathbf{x}, \tag{1.8}$$

where $R = \mathbf{M}^{-1}(T)$.

In view of the preceding discussion no loss of generality results in assuming, whenever necessary, that the distribution is given by (1.1).

We shall list, briefly review and discuss a number of important distributional problems, together with some applications, which are formally reduceable to integrals of the form (1.1). In the first few illustrations, the regions R constitute relatively simple geometrical entities, such as half-spaces, hyperspheres, hypercones and hypercylinders, for which the statistical applications are both classic and familiar, but in later illustrations more complex bodies, such as ellipsoids, simplices and polyhederal cones, are considered. In particular, the last named case of polyhederal cones, corresponding to the difficult and important problem of the multivariate normal integral, and more especially the bivariate normal integral (when the dimensionality of the polyhedral cone is 2), will be investigated in some detail.

Integrals of the form (1.1) are rarely capable of being expressed in closed form using well-known functions. Nevertheless, it is hoped that the current presentation will provide a unifying thread and thereby help to stimulate further research. In the sequel a quite powerful method, referred to as the "method of sections," will frequently be used to deal with the integrals. This consists in dividing up the region R by means of a series of parallel and adjoining $(N - 1)$-flats and in the exploitation of the following fundamental property of the spherical normal distribution of dimensionality N: The conditional probability distribution in any linear subspace of dimensionality $N - k$ $(k = 1, 2, \cdots, N - 1)$

is itself spherical normal with dimensionality $N - k$ and with variance in any direction equal to the variance of the original N-dimensional distribution. Let O be the center of distribution, P any point in R and M the foot of the perpendicular from O to the flat through P. Further, let $OP = r$, $OM = \xi$, $PM = \eta$, with $r^2 = \xi^2 + \eta^2$. Then the p.d.f. at P is

$$(2\pi)^{-\frac{1}{2}N}e^{-\frac{1}{2}r^2} = (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}\xi^2} \times (2\pi)^{-\frac{1}{2}(N-1)}e^{-\frac{1}{2}\eta^2},$$

and the distribution in the flat through P is spherical normal with dimensionality $N - 1$. It follows that the probability content of the infinitesimal region intercepted by R between two parallel flats distant ξ and $\xi + d\xi$ from O is of the form

$$(1.9) \qquad (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}\xi^2}\, d\xi Q(\xi; R),$$

where $Q(\xi, R)$ is itself obtained by evaluating an integral of the form (1.1), with N replaced by $N - 1$. Consequently,

$$(1.10) \qquad P(R) = \int_{\xi_0}^{\xi_1} (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}\xi^2}Q(\xi; R)\, d\xi,$$

where ξ_0 and ξ_1 are the distances of the bounding flats to R from O. If, further, the section of each cutting flat is a region of the same geometrical type as R (e.g. R an ellipsoid and the section an ellipsoid), with the center M of the $(N - 1)$-dimensional spherical distribution in the flat bearing the same geometrical relationship with respect to the $(N - 1)$-dimensional figure as does O with respect to R (e.g. both O and M are centers of ellipsoids), then (1.10) becomes an integral recurrence relationship (see Sections 7 and 8).

2. Probability content of a half-space. The probability content of the infinite parallel slab R defined by $p_1 \leq \sum_1^N a_i x_i \leq p_2$ is given directly by the method of sections as

$$(2.1) \qquad (2\pi)^{-\frac{1}{2}} \int_{p_1/(\Sigma a_i^2)^{\frac{1}{2}}}^{p_2/(\Sigma a_i^2)^{\frac{1}{2}}} e^{-\frac{1}{2}\xi^2}\, d\xi.$$

Here the flats dividing R are taken parallel to the bounding flats and $Q(\xi; R) = 1$, $\xi_0 = p_1/(\sum a_i^2)^{\frac{1}{2}}$, $\xi_1 = p_2/(\sum a_i^2)^{\frac{1}{2}}$. In particular, for the lower half-space $p_1 = -\infty$ and (2.1) becomes

$$(2.2) \qquad (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{p_2/(\Sigma a_i^2)^{\frac{1}{2}}} e^{-\frac{1}{2}\xi^2}\, d\xi.$$

These results are, of course, a reflection of the fact that $\sum a_i x_i$ is distributed normally with zero mean and variance $\sum a_i^2$.

3. Probability contents of centrally and non-centrally located hyperspheres. Historically, the central χ^2 distribution was one of the first directly entailing probability contents of regions in N-space when the density is spherical normal.[2]

[2] A geometrical derivation of the χ^2 distribution *for 3 degrees of freedom* is implicit in Maxwell's great work [2] concerning the energy distribution of gas molecules. Each of three orthogonal components of velocity have identical and independent normal distributions with zero mean, and the energy, suitably standardized, is a χ^2 with 3 degrees of freedom.

For the central χ^2 distribution the region in question is a sphere whose center coincides with the center of the distribution, while for the non-central distribution the region is a sphere whose center is non-coincident with the center of the distribution.

Let χ_N^2 and $\chi_{N;\kappa}^2$ refer generically to a variate distributed as χ^2 with N degrees of freedom and a non-central χ^2 variate with N degrees of freedom and non-centrality parameter κ, respectively ($\chi_{N;0}^2 \equiv \chi_N^2$). The latter variate is defined by $\chi_{N;\kappa}^2 = \sum_1^N (x_i - \kappa_i)^2$, where the x_i are independent normal variables with zero means and unit variances, and $\kappa^2 = \sum_1^N \kappa_i^2$. Further, denote the distribution functions of these two variates by $F_N(a^2)$ and $G_{N;\kappa}(a^2)$. Correspondingly, lower case letters shall denote the p.d.f.'s. Then

$$
\begin{aligned}
F_N(a^2) = P(\chi_N^2 \leqq a^2) &= \int \cdots \int_{R:\sum_1^N x_i^2 \leqq a^2} (2\pi)^{-\frac{1}{2}N} \exp\left(-\tfrac{1}{2}\sum_1^N x_i^2\right) dx_1 \cdots dx_N \\
&= \int \cdots \int_R (2\pi)^{-\frac{1}{2}N} \exp(-\tfrac{1}{2}r^2) r^{N-1}\, dr\, d\omega \\
&= S_N(1) \int_0^a (2\pi)^{-\frac{1}{2}N} \exp(-\tfrac{1}{2}r^2) r^{N-1}\, dr \\
&= \left(2^{\frac{1}{2}N}\Gamma\left(\frac{N}{2}\right)\right)^{-1} \int_0^{a^2} \exp(-\tfrac{1}{2}r^2)(r^2)^{\frac{1}{2}N-1}\, dr^2,
\end{aligned}
\tag{3.1}
$$

where $d\omega$ is the solid angle subtended at the center of the distribution by an infinitesimal volume element and $S_N(c)$ is the surface-content of a hypersphere of N dimensions with radius c,

$$S_N(c) = 2\pi^{\frac{1}{2}N} c^{N-1}/\Gamma(N/2). \tag{3.2}$$

This gives the usual Incomplete Gamma Function for the distribution function. On differentiating with respect to a^2,

$$f_N(a^2) = [2^{\frac{1}{2}N}\Gamma(N/2)]^{-1} \exp(-\tfrac{1}{2}a^2)(a^2)^{\frac{1}{2}N-1}. \tag{3.3}$$

Pedagogically, perhaps a more useful geometrical derivation is to "slice" up the sphere into infinitesimal thin slices by a set of parallel planes. This corresponds to a proof by induction (cf. [3], pp. 247–8). Let x denote the distance of a typical slice from the center of the sphere. Then

$$
\begin{aligned}
F_N(a^2) &= \int \cdots \int_{R:\sum_1^N x_i^2 \leqq a^2} (2\pi)^{-\frac{1}{2}N} \exp\left(-\frac{1}{2}\sum_1^N x_i^2\right) dx_1 \cdots dx_N \\
&= \int_{-a}^{a} (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}x^2)\cdot F_{N-1}(a^2 - x^2)\, dx,
\end{aligned}
\tag{3.4}
$$

on noting that the density at a point on the "x-slice", which intersects the given sphere in a sphere of dimensionality $N - 1$ and radius $(a^2 - x^2)^{\frac{1}{2}}$, distant y

from the center of the latter sphere, is $(2\pi)^{-N/2}\exp[-\frac{1}{2}(x^2+y^2)]$. On differentiating with respect to a^2,

$$
\begin{aligned}
f_N(a^2) &= \int_{-a}^{a}(2\pi)^{-\frac{1}{2}}\exp(-\tfrac{1}{2}x^2)f_{N-1}(a^2-x^2)\,dx\\
&= \int_{-a}^{a}(2\pi)^{-\frac{1}{2}}\exp(-\tfrac{1}{2}x^2)\\
&\qquad\left[2^{\frac{1}{2}(N-1)}\Gamma\left(\frac{N-1}{2}\right)\right]^{-1}\exp[-\tfrac{1}{2}(a^2-x^2)](a^2-x^2)^{\frac{1}{2}(N-3)}\,dx\\
&= \exp(-\tfrac{1}{2}a^2)\left[(2\pi)^{\frac{1}{2}}2^{\frac{1}{2}(N-1)}\Gamma\left(\frac{N-1}{2}\right)\right]^{-1}\int_{-a}^{a}(a^2-x^2)^{\frac{1}{2}(N-3)}\,dx\\
&= \left(2^{\frac{1}{2}N}\Gamma\left(\frac{N}{2}\right)\right)^{-1}\exp(-\tfrac{1}{2}a^2)(a^2)^{\frac{1}{2}N-1}.
\end{aligned}
$$

For the non-central χ^2 distribution, we require the distribution of $\sum_1^N(x_i-\kappa_i)^2$, where the x_i are mutually independent standardized normal random variables. Let O be the center of the distribution which is taken as before to be the origin of coordinates and K the point $(\kappa_1, \kappa_2, \cdots, \kappa_N)$. Let P be any point with coordinates $(x_1, x_2, \cdots, x_N)$, let $OK = \kappa$, $\kappa = (\kappa_1^2+\kappa_2^2+\cdots+\kappa_N^2)^{\frac{1}{2}}$, $KP = \xi$, and let the angle between KP and the line OK, produced in the sense O to K, be θ. Then

$$\sum_1^N x_i^2 = OP^2 = \kappa^2+\xi^2+2\kappa\xi\cos\theta,$$

and

$$
\begin{aligned}
G_{N;\kappa}(a^2) &= P\left\{\sum_1^N(x_i-\kappa_i)^2 \le a^2\right\}\\
&= \underset{R:\sum_1^N(x_i-\kappa_i)^2\le a^2}{\int\cdots\int}(2\pi)^{-\frac{1}{2}N}\exp\left(-\tfrac{1}{2}\sum_1^N x_i^2\right)dx_1\cdots dx_N\\
&= \int_0^a\int_0^\pi(2\pi)^{-\frac{1}{2}N}\exp[-\tfrac{1}{2}(\kappa^2+\xi^2+2\kappa\xi\cos\theta)]\xi^{N-1}\,d\xi\,d\omega\\
&= (2\pi)^{-\frac{1}{2}N}\exp(-\tfrac{1}{2}\kappa^2)\int_0^a\int_0^\pi\exp(-\tfrac{1}{2}\xi^2-\kappa\xi\cos\theta)\xi^{N-1}\,d\xi\,d\omega.
\end{aligned}
$$

Now

$$\int_0^\pi\exp(-\kappa\xi\cos\theta)\,d\omega = 2\pi^{\frac{1}{2}N}(\tfrac{1}{2}\kappa\xi)^{-(\frac{1}{2}N-1)}I_{\frac{1}{2}N-1}(\kappa\xi),$$

where $I_n(z) = i^{-n}J_n(iz)$ is the Bessel function of the first kind with purely imaginary argument. This follows directly by dividing up the surface of the hypersphere into annuli $d\theta$, the content of such an annulus being $S_{N-1}(\sin\theta)\,d\theta$.

Thus,

$$\int_0^\pi \exp(-\kappa\xi\cos\theta)\,d\omega = \int_0^\pi \exp(-\kappa\xi\cos\theta)S_{N-1}(\sin\theta)\,d\theta$$

$$= \left[2\pi^{\frac{1}{2}(N-1)} \Big/ \Gamma\left(\frac{N-1}{2}\right)\right]\int_0^\pi \exp(-\kappa\xi\cos\theta)\sin^{N-2}\theta\,d\theta,$$

and this integral is related to the Bessel function

$$I_n(z) = [\sqrt{\pi}\,\Gamma(n+\tfrac{1}{2})]^{-1}(\tfrac{1}{2}z)^n\int_{-1}^{1}\exp(\pm zv)(1-v^2)^{n-\frac{1}{2}}\,dv \quad (R(n+\tfrac{1}{2})>0),$$

by setting $v = \cos\theta$. Alternatively, $\exp(-\kappa\xi\cos\theta)$ may be expanded as a power series in $\cos\theta$ after which integration is effected term by term. Hence, finally,

$$\begin{aligned} G_{N;\kappa}(a^2) &= (2\pi)^{-\frac{1}{2}N}\exp(-\tfrac{1}{2}\kappa^2) \\ &\quad\cdot\int_0^a \exp(-\tfrac{1}{2}\xi^2)\xi^{N-1}\cdot 2\pi^{\frac{1}{2}N}(\tfrac{1}{2}\kappa\xi)^{-(\frac{1}{2}N-1)}I_{\frac{1}{2}N-1}(\kappa\xi)\,d\xi \end{aligned} \tag{3.5}$$

and

$$\begin{aligned} g_{N;\kappa}(a^2) &= \tfrac{1}{2}\kappa^{-(\frac{1}{2}N-1)}\exp(-\tfrac{1}{2}\kappa^2)a^{\frac{1}{2}N-1}\exp(-\tfrac{1}{2}a^2)I_{\frac{1}{2}N-1}(\kappa a) \\ &= 2^{-\frac{1}{2}N}\exp(-\tfrac{1}{2}\kappa^2)(a^2)^{\frac{1}{2}N-1}\exp(-\tfrac{1}{2}a^2)\sum_{r=0}^{\infty}[1/\Gamma(\tfrac{1}{2}N+r)] \\ &\quad\cdot[(\tfrac{1}{2}\kappa a)^{2r}/r!]. \end{aligned} \tag{3.6}$$

The above geometrical derivation seems to have been used first, in essence, by Patnaik [4].

An alternative and simpler geometrical method consists once again in dividing the sphere R by a set of parallel hyperplanes. Take these to be perpendicular to the line OK and let x be the distance of a typical plane from K. Then

$$\begin{aligned} G_{N;\kappa}(a^2) &= \int\cdots\int_{R:\sum_1^N(x_i-\kappa_i)^2\le a^2}(2\pi)^{-\frac{1}{2}N}\exp\left(-\frac{1}{2}\sum_1^N x_i^2\right)dx_1\cdots dx_N \\ &= \int_{-a}^{a}(2\pi)^{-\frac{1}{2}}\exp[-\tfrac{1}{2}(\kappa+x)^2]F_{N-1}(a^2-x^2)\,dx. \end{aligned} \tag{3.7}$$

Hence,

$$\begin{aligned} g_{N;\kappa}(a^2) &= \exp(-\tfrac{1}{2}\kappa^2)\int_{-a}^{a}\exp(-\kappa x)\cdot(2\pi)^{-\frac{1}{2}}\exp(-\tfrac{1}{2}x^2)f_{N-1}(a^2-x^2)\,dx \\ &= \exp(-\tfrac{1}{2}\kappa^2)\exp(-\tfrac{1}{2}a^2)\Big/\left[(2\pi)^{\frac{1}{2}}2^{\frac{1}{2}(N-1)}\Gamma\left(\frac{N-1}{2}\right)\right] \\ &\quad\cdot\int_{-a}^{a}\exp(-\kappa x)(a^2-x^2)^{\frac{1}{2}(N-3)}\,dx \\ &= \tfrac{1}{2}\kappa^{-(\frac{1}{2}N-1)}\exp(-\tfrac{1}{2}\kappa^2)a^{\frac{1}{2}N-1}\exp(-\tfrac{1}{2}a^2)I_{\frac{1}{2}N-1}(\kappa a), \end{aligned}$$

after substitution for $f_{N-1}(a^2 - x^2)$ from (3.3) and using the above integral formula for $I_n(z)$.

Some exact values for the probability integral of the non-central χ^2 are available in [5] and [6]. A more extensive set of values is provided in Fix's tables [7] designed to yield the power function of χ^2. For studies in connexion with suitable approximations to the non-central χ^2 distribution, reference is made to [4], [8] and [9]. Finally, various tables of the non-central χ^2 distribution for the special case of two degrees of freedom have become available in recent years for application in ballistic problems ([10], [11], [12], [13]).

4. Probability content of a symmetrically and asymmetrically located hyperspherical cone. Consider a hyperspherical half-cone R with vertex at the center of the spherical normal distribution. Let the angle between the axis of the cone and a generator be θ. The probability content $P(R)$ of the cone is given by its relative solid angle, i.e., the ratio of the surface-content of the region, a cap, on a unit sphere with center at the vertex of the cone which is demarcated by the cone to the surface-content of the entire sphere. Hence, by division of the cap into a set of annuli with radii sin θ',

$$
\begin{aligned}
P(R) &= \int_0^\theta S_{N-1}(\sin\theta')\,d\theta'/S_N(1) \\
&= \Gamma\left(\frac{N}{2}\right)\Big/\left[\sqrt{\pi}\,\Gamma\left(\frac{N-1}{2}\right)\right]\int_0^\theta \sin^{N-2}\theta'\,d\theta' \qquad (4.1)\\
&= \mathscr{I}_{\sin^2\theta}\left(\frac{N-1}{2}, \tfrac{1}{2}\right),
\end{aligned}
$$

where $\mathscr{I}$ denotes the Incomplete Beta Function *Ratio*.

Define

$$t_{N-1} = \xi/(\chi^2_{N-1}/(N-1))^{\frac{1}{2}}, \tag{4.2}$$

where ξ is a normal variate with zero mean and unit variance distributed independently of χ^2_{N-1}, a χ^2 variate with $N - 1$ degrees of freedom. The variate t_{N-1} may be expressed in the form

$$t_{N-1} = x_1\Big/\left(\sum_2^N x_i^2/(N-1)\right)^{\frac{1}{2}}, \tag{4.3}$$

where the x_i $(i = 1, 2, \cdots, N)$ are independent normal variates, each with zero mean and unit variance. The region $t_{N-1} \geq c(c \geq 0)$ defines a half-cone with vertex at the center of the distribution of the x_i and with axis oriented along the x_1-axis. The angle between the axis and a generator is $\theta = \text{arc cot}\,[c/(N-1)^{\frac{1}{2}}]$. The distribution function of t_{N-1} is given by (4.1), where here

$$\sin^2\theta = 1/[1 + c^2/(N-1)]. \tag{4.4}$$

The density function $-\partial P/\partial c$ is

$$g_{N-1}(c) = \left[(N-1)^{\frac{1}{2}}B\left(\frac{N-1}{2}, \tfrac{1}{2}\right)\right]^{-1}\left(1 + \frac{c^2}{N-1}\right)^{-\frac{1}{2}N}. \tag{4.5}$$

The simplest application of the t-distribution relates to the "Studentized" mean of a normal sample. The geometrical derivation of the latter quantity is well-known (see e.g. [3], pp. 239–40). The relevant hyperspherical cone in this instance has its axis along the line $x_1 = x_2 = \cdots = x_N$ equally inclined to the coordinate planes. The above argument implies, however, that the probability content of a hyperspherical cone of given angle and with vertex at the center of a spherical normal distribution is independent of its orientation.

Consider next a hyperspherical cone whose vertex does not coincide with the center of a given spherical normal distribution but whose axis passes through the latter point. As before, let the angle between the axis and a generator be θ. The probability content $P(R)$ of the cone may be obtained by considering sections perpendicular to the generator. Each such section is a hypersphere and the surfaces of equal density in the flat forming the section are hyperspheres. Hence[3],

$$P(R) = \int_{\lambda}^{\infty} (2\pi)^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}x^2\right) F_{N-1}[(x-\lambda)^2 \tan^2\theta]\, dx, \tag{4.6}$$

where λ is the distance of the vertex from the center of the distribution, and $F_{N-1}(\cdot)$ is defined in (3.1).

Define

$$t_{N-1;\lambda} = [\xi - \lambda]/[\chi^2_{N-1}/(N-1)]^{\frac{1}{2}}, \tag{4.7}$$

where ξ is a normal variate with zero mean and unit variance distributed independently of χ^2_{N-1}, a χ^2 variate with $N-1$ degrees of freedom ($t_{N-1} \equiv t_{N-1;0}$). The variate $t_{N-1;\lambda}$ may be expressed in the form

$$t_{N-1;\lambda} = [x_1 - \lambda] \Big/ \left[\sum_{2}^{N-1} x_i^2/(N-1)\right]^{\frac{1}{2}}, \tag{4.8}$$

where the $x_i (i = 1, 2, \cdots, N)$ are independent normal variates each with zero mean and unit variance. The region $t_{N-1;\lambda} \geq c (c \geq 0)$ defines a half-cone with vertex distant λ from the center of the distribution of the x_i. The axis of the cone is oriented along the x_1-axis and the angle between the latter and a generator is arc cot $[c/(N-1)^{\frac{1}{2}}]$. The distribution function of $t_{N-1;\lambda}$ is given by (4.6) with $\theta = \text{arc cot}\,[c/(N-1)^{\frac{1}{2}}]$. Setting $y = x - \lambda$ and differentiating with respect to y, the density $-\partial P/\partial c$ of $t_{N-1;\lambda}$ at c is obtained immediately as

$$\begin{aligned} g_{N-1;\lambda}(c) &= g_{N-1}(c) \exp\left(-\tfrac{1}{2}\lambda^2\right)\left[\Gamma\left(\frac{N}{2}\right)\right]^{-1} \\ &\quad \cdot \int_0^{\infty} z^{\frac{1}{2}N-1} \exp\left[-z - \lambda\sqrt{2z}\cos\theta\right] dz, \end{aligned} \tag{4.9}$$

where $y^2 \sec^2\theta/2$ has been replaced by z. This density function may be expressed in terms of the Hh function [14], defined by

$$Hh_m(y) = \int_0^{\infty} [x^m/m!] \exp\left[-\tfrac{1}{2}(x+y)^2\right] dx,$$

[3] This argument incidentally provides an alternative basis for the determination of the distribution and density functions of t_{N-1}, i.e., when $\lambda = 0$.

as follows:

$$g_{N-1;\lambda}(c) = g_{N-1}(c)\,\Gamma(N)\left[2^{\frac{1}{2}N-1}\Gamma\left(\frac{N}{2}\right)\right]^{-1} \cdot \exp\left(-\tfrac{1}{2}\lambda^2 \sec^2\theta\right) Hh_{N-1}(\lambda\cos\theta), \tag{4.10}$$

θ being given by (4.4). Alternatively, term by term integration yields

$$g_{N-1;\lambda}(c) = \frac{\exp(-\frac{1}{2}\lambda^2)}{\Gamma\left(\frac{N-1}{2}\right)((N-1)\pi)^{\frac{1}{2}}} \sum_{r=0}^{\infty} \frac{(-\lambda\sqrt{2})^r}{r!} \cdot \Gamma\left(\frac{N+r}{2}\right)\left(\frac{c^2}{N-1}\right)^{\frac{1}{2}r}\left(1+\frac{c^2}{N-1}\right)^{-\frac{1}{2}(N+r)}. \tag{4.11}$$

The "tail-end area," obtained after term by term integration in (4.11), yields

$$P(t_{N-1;\lambda} \geqq c) = \frac{\exp(-\frac{1}{2}\lambda^2)}{2\sqrt{\pi}} \sum_{r=0}^{\infty} \Gamma[\tfrac{1}{2}(r+1)] \, \mathcal{I}_{\sin^2\theta}\left(\frac{N-1}{2}, \frac{r+1}{2}\right) \frac{(-\lambda\sqrt{2})^r}{r!} \qquad (c \geqq 0) \tag{4.12}$$

(cf. [15]). Tables of the non-central t-distribution have been provided by Neyman and Tokarska [16], Johnson and Welch [17] and, more recently, by Resnikoff and Lieberman [18] together with applications. The simplest application relates to the power of the test based on the Studentized mean-statistic from a normal sample. The axis of the relevant cone is then along the line $x_1 = x_2 = \cdots = x_N$. The above argument implies, however, that the probability content of a hyperspherical cone of given angle and with axis passing through the center of a spherical normal distribution is independent of its orientation provided that the distance between the vertex and the center of the distribution remains fixed.

5. Probability content of a region bounded by a variety of revolution of dimensionality $N - 1$ and of species p. Denote a hyperspherical surface (manifold) of dimensionality m by S_m. Then a variety of revolution S_{N-1} of dimensionality $N - 1$ and of species p is defined by the rotation of a S_{N-p-1}, imbedded in a $(N-p)$-flat A_{N-p} (linear space of dimensionality $N - p$), round a $(N - p - 1)$-flat Λ_{N-p-1} in A_{N-p} as axis (see e.g. Sommerville [19], pp. 137–8). The axial plane of revolution Λ_{N-p-1} may be regarded as defined by $N - p$ fixed points in a $(N-1)$-flat A_{N-1} imbedded in N-space (Λ_{N-p-1} is a linear subspace of A_{N-1}) which has therefore p degrees of freedom and can rotate about Λ_{N-p-1} in such a way that each point of S_{N-p-1} generates the surface of a hypersphere with dimensionality $p + 1$, the latter surface itself being of dimensionality p. The center of the hypersphere is determined by the foot of the perpendicular to Λ_{N-p-1} from the given point.

If the equation of the generating surface S_{N-p-1}, referred to $N - p$ rectangular

axes in A_{N-p} of which $N-p-1$, designated the x_1-axis, $\cdots$, x_{N-p-1}-axis, are in Λ_{N-p-1}, is given by

$$x_{N-p}^2 = \phi(x_1, x_2, \cdots, x_{N-p-1}), \tag{5.1}$$

then the equation of the generated surface S_{N-1} is

$$x_{N-p}^2 + x_{N-p+1}^2 + \cdots + x_N^2 = \phi(x_1, x_2, \cdots, x_{N-p-1}), \tag{5.2}$$

since the expression on the left of equ. (5.2) represents the squared perpendicular distance of a point in S_{N-p-1}, when the latter is in a rotated position, from the axis Λ_{N-p-1}.

We shall now determine the probability content of the region R in N-space obtained by replacing the quality sign in equ. (5.2) by the sign $\leqq$, under the assumption that the distribution of the $x_i (i = 1, 2, \cdots, N)$ is governed by (1.1).

Let O be the center of the distribution, P the point $(x_1, x_2, \cdots, x_N)$ on S_{N-p-1} and P' the point $(x_1, \cdots, x_{N-p-1}, 0, \cdots, 0)$, P' being the foot of the perpendicular to Λ_{N-p-1} from P. The locus of P on rotation of S_{N-p-1} is a hypersphere with radius $\phi(x_1, x_2, \cdots, x_{N-p-1})$. Consider the infinitesimal element of R which projects into the element $d\Lambda_{N-p-1}$ located in Λ_{N-p-1} around P'. Since the density at any point is

$$(2\pi)^{-\frac{1}{2}N} \exp\left(-\tfrac{1}{2}\sum_1^N x_i^2\right) = (2\pi)^{-\frac{1}{2}(N-p-1)} \exp\left(-\tfrac{1}{2}\sum_1^{N-p-1} x_i^2\right) \cdot (2\pi)^{-\frac{1}{2}(p+1)} \exp\left(-\tfrac{1}{2}\sum_{N-p}^N x_i^2\right),$$

the distribution in the p-flat containing the locus of P is spherical normal with center P'. Hence, by (3.1), the probability content of the element is

$$(2\pi)^{-\frac{1}{2}(N-p-1)} \exp\left(-\tfrac{1}{2}\sum_1^{N-p-1} x_i^2\right) F_{p+1}(\phi(x_1, x_2, \cdots, x_{N-p-1}))\, d\Lambda_{N-p-1},$$

on recalling that $PP'^2 = \phi(x_1, x_2, \cdots, x_{N-p-1})$, where $F_{p+1}(\cdot)$ denotes the distribution function of a chi-square with $p+1$ degrees of freedom (equ. (3.1)). The required probability content is then

$$\int \cdots \int (2\pi)^{-\frac{1}{2}(N-p-1)} \exp\left(-\tfrac{1}{2}\sum_1^{N-p-1} x_i^2\right) \cdot F_{p+1}(\phi(x_1, x_2, \cdots, x_{N-p-1}))\, d\Lambda_{N-p-1}. \tag{5.3}$$

Consider in particular the case

$$\phi(x_1, \cdots, x_{N-p-1}) = [(p+1)/k(N-p-1)](x_1^2 + \cdots + x_{N-p-1}^2) \tag{5.4}$$

when the region R becomes

$$[(x_1^2 + \cdots + x_{N-p-1}^2)/(N-p-1)]/[(x_{N-p}^2 + \cdots + x_N^2)/(p+1)] \geqq k. \tag{5.5}$$

Equation (5.1) now represents a hyperspherical conical surface in Λ_{N-p}, while

equ. (5.2) represents the surface obtained by rotation round Λ_{N-p-1}. The generated surface is characterized by the property that the radius vector OP is inclined at a constant angle arc cot $(k(N - p - 1)/(p + 1))^{\frac{1}{2}}$ to the linear space Λ_{N-p-1}. Furthermore, in (5.3) $d\Lambda_{N-p-1}$ may conveniently be chosen as the annulus between two concentric hyperspherical surfaces in the Λ_{N-p-1}-subspace of radii ξ and $\xi + d\xi(\xi^2 = \sum_1^{N-p-1} x_i^2 = OP'^2)$, and the probability content of R is then

$$(5.6)\qquad \int_0^\infty (2\pi)^{-\frac{1}{2}(N-p-1)} \exp\left(-\tfrac{1}{2}\xi^2\right) \frac{2\pi^{\frac{1}{2}(N-p-1)}\xi^{N-p-2}}{\Gamma\left(\frac{N-p-1}{2}\right)} F_{p+1}\left[\frac{(p+1)\xi^2}{k(N-p-1)}\right] d\xi,$$

on using equ. (3.2) to substitute for $d\Lambda_{N-p-1}$.

The expression (5.6) represents $1 - P_{N-p-1,p+1}(k)$, where $P_{N-p-1,p+1}(\cdot)$ denotes the distribution function of an F-variate, $F_{N-p-1,p+1}$, with $N - p - 1$ and $p + 1$ degrees of freedom. The density function of the F-variate at the point k is

$$-\partial P_{N-p-1,p+1}/\partial k = \frac{2^{-\frac{1}{2}(N-2)}(p+1)^{\frac{1}{2}(p+1)}}{\Gamma\left(\frac{N-p-1}{2}\right)\Gamma\left(\frac{p+1}{2}\right)(N-p-1)^{\frac{1}{2}(p+1)}}$$

$$(5.7)\qquad \cdot \frac{1}{k^{\frac{1}{2}(p+3)}} \int_0^\infty \xi^{N-1} \exp\left\{-\tfrac{1}{2}\left[1 + \frac{p+1}{(N-p-1)k}\right]\xi^2\right\} d\xi$$

$$= \frac{1}{B\left(\frac{N-p-1}{2}, \frac{p+1}{2}\right)} (N-p-1)^{\frac{1}{2}(N-p-1)}(p+1)^{\frac{1}{2}(p+1)} \cdot \frac{k^{\frac{1}{2}(N-p-1)-1}}{[(N-p-1)k + (p+1)]^{\frac{1}{2}N}},$$

after some reduction.

The case discussed in Section 4 corresponds to $p = N - 2$.

6. Probability contents of symmetrically and asymmetrically located hyperspherical cylinders. A hyperspherical cylinder in N-space is one such that the intersection with the cylinder of a $(N - 1)$-flat perpendicular to the axis of the cylinder is a hypersphere.

There are two distinct cases to consider:

(a) The axis of the cylinder passes through the center of the distribution.

(b) The axis of the cylinder does not pass through the center of the distribution.

The probability content may in both cases be readily evaluated by taking sections perpendicular to the axis. Let a be the radii of the cylinders in both (a) and (b), and let λ be the distance between the axis of the cylinder and the center of the distribution in (b). The probability contents of elements formed by adjoining parallel $(N - 1)$-flats distant x and $x + dx$ from the center of the distribution perpendicular to the axis of the cylinder is seen directly to be

$(2\pi)^{-\frac{1}{2}} \exp(-\frac{1}{2}x^2)\, dx F_{N-1}(a^2)$ and $(2\pi)^{-\frac{1}{2}} \exp(-\frac{1}{2}x^2)\, dx G_{N-1;\lambda}(a^2)$, respectively. Hence, by integration of x over $(-\infty, \infty)$, the probability content of the cylinder in case (a) is $F_{N-1}(a^2)$ and that of the cylinder in case (b) is $G_{N-1;\lambda}(a^2)$. A particularly simple application of (a) relates to the distribution of the sample variance in normal samples when the cylinder in question has its axis along the line $x_1 = x_2 = \cdots = x_N$ ([3], p. 238).

7. Probability content of a centrally situated ellipsoid. The problem treated in this section is equivalent to that of finding the distribution of the weighted sum of squares of mutually independent standardized normal variates. Formally, we require

$$(7.1) \qquad F_{N;a_1,\cdots,a_N}(t) = P\left(\sum_1^N a_i x_i^2 \leqq t\right), \qquad a_i \geqq 0, \qquad \sum_1^N a_i = 1,$$

where the x_i are the variates referred to. The center of the ellipsoid coincides with the center of the distribution and the lengths of the semi-axes are $(t/a_i)^{\frac{1}{2}} (i = 1, 2, \cdots, N)$. The axes are oriented along the coordinate axes.

The distribution of $\sum a_i x_i^2$ has been discussed by Bhattacharya [20], Robbins [21], Robbins and Pitman [22], Hotelling [23], Gurland [24], [25], Pachares [26] and by Grad and Solomon [27]. The latter authors have tabulated $F_{N;a_1,\cdots,a_N}(t)$ for $N = 2, 3$, and for various selected sets of (a_1, a_2) and (a_1, a_2, a_3). We shall here obtain an integral recurrence relationship, based on the method of sections used previously in this paper, which should enable a systematic extension to be made of the available tables to values of $N > 3$, at least for moderate N[4], as well as of the tables for $N = 2, 3$. The following additional remarks are pertinent:

(i) There is no loss of generality in assuming $\sum a_i = 1$, since this can always be achieved by suitable standardization. However, the weights a_i are all non-negative.

(ii) The important statistical problem of the distribution of the weighted sum of independent χ^2 variates may be considered as a special case of our problem. Specifically, if $y = \sum_1^k c_i u_i$ where the u_i are independent χ^2 variates with n_i degrees of freedom, $\sum_1^k n_i = N$, then since N independent standardized normal variates, $x_1, x_2, \cdots, x_N$, may be introduced so that

$$(7.2) \qquad u_i = \sum_{j=1}^{n_i} x^2_{n_1+n_2+\cdots+n_{i-1}+j} \qquad (i = 1, 2, \cdots, k),$$

y may be expressed in terms of the x_i in the form

$$y = \sum_{i=1}^{k} \sum_{j=1}^{n_i} c_i x^2_{n_1+n_2+\cdots+n_{i-1}+j}$$

$$(7.3) \qquad = \sum_{\alpha=1}^{N} a_\alpha x_\alpha^2, \qquad a_\alpha = c_i \quad \text{for} \quad \alpha = n_1 + n_2 + \cdots + n_{i-1} + j$$
$$(j = 1, 2, \cdots, n_i).$$

[4] Extension of the Grad and Solomon tables for $N = 2$ and $N = 3$ has now been effected by Professor H. Solomon and the present author with the aid of (7.10). It is hoped to publish the extended tables shortly.

(iii) The problem of determining the distribution of a definite positive quadratic function of N variables when the latter are distributed as in a non-degenerate multivariate normal distribution reduces easily to our problem. Geometrically, the probability content of a given ellipsoid is required when the surfaces of constant density of the normal distribution are those of homothetic ellipsoids. A rotation of the coordinate axes and subsequent scaling converts the latter ellipsoids into spheres, while the given ellipsoid will in general remain an ellipsoid under these two transformations. Finally, a rotation of the new coordinate axes to bring them into coincidence with the axes of the given ellipsoid is effected.

Formally, one desires to evaluate the quantity

$$(7.4) \qquad P(\mathbf{x}'\mathbf{A}\mathbf{x} \leqq c^2) = (2\pi)^{-\frac{1}{2}N} \mid \mathbf{V} \mid^{-\frac{1}{2}} \int_{\mathbf{x}'\mathbf{A}\mathbf{x} \leqq c^2} \exp(-\tfrac{1}{2}\mathbf{x}'\mathbf{V}^{-1}\mathbf{x})\, d\mathbf{x},$$

in which $\mathbf{A}$ and $\mathbf{V}$ are each of rank N. Set $\mathbf{x} = \mathbf{LRy}$, where $\mathbf{V}$ is decomposed by triangular resolution (as in the introductory section) in the form $\mathbf{V} = \mathbf{LL}'$, $\mathbf{L}$ being a lower $N \times N$ triangular matrix, while $\mathbf{R}$ is the orthogonal matrix of the characteristic vectors of the matrix $\mathbf{L}'\mathbf{AL}$. Then, after substitution,

$$(7.5) \qquad P(\mathbf{x}'\mathbf{A}\mathbf{x} \leqq c^2) = (2\pi)^{-\frac{1}{2}N} \int_{\mathbf{y}'\mathbf{\Lambda}\mathbf{y} \leqq c^2} \exp(-\tfrac{1}{2}\mathbf{y}'\mathbf{y})\, d\mathbf{y},$$

where the diagonal matrix $\mathbf{\Lambda}$ is given by $\mathbf{\Lambda} = \mathbf{R}'\mathbf{LALR}$.

If the diagonal elements of $\mathbf{\Lambda}$ are denoted by $\lambda_1, \lambda_2, \cdots, \lambda_N$, the characteristic numbers of $\mathbf{L}'\mathbf{AL}(\lambda_i > 0, i = 1, 2, \cdots, N)$, equation (7.5) is equivalent to

$$P(\mathbf{x}'\mathbf{A}\mathbf{x} \leqq c^2) = P\left(\sum_1^N \lambda_i y_i^2 \leqq c^2\right),$$

in which the y_i are mutually independent normal variates with zero means and unit variances. This establishes the equivalence of the problem dealt with in this subsection with that of equation (7.1).

(iv) A similar argument is applicable to the situations in which $\mathbf{A}$ is semi-definite positive. Here one wishes to evaluate the probability content of an elliptic cylinder under spherical normal distributions. The latter is clearly equal to the probability content of the ellipsoid, relating to an appropriately chosen subspace, obtained by projection into the latter subspace. The dimensionality of the subspace is equal to the rank of $\mathbf{A}$. A case in point is the mean square successive difference $\delta^2_{(1)}$, defined by

$$(7.6) \qquad 2(N-1)\delta^2_{(1)} = \mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{N-1} (x_{i+1} - x_i)^2$$

([28], [29]), for which $\mathbf{A}$ is the continuant

$$\begin{bmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & \cdot & \cdot & \cdot & & \\ & & & \cdot & \cdot & \cdot & \\ & & & & \cdot & \cdot & \cdot \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{bmatrix},$$

empty spaces denoting zeros. The mean square successive difference has been proposed as a suitable estimator of variability when a secular trend in the mean is suspected. The inequality $\mathbf{x}'\mathbf{A}\mathbf{x} \leq c^2$ defines the interior and boundary of an elliptic cylinder with axis $x_1 = x_2 = \cdots = x_N$ equally inclined to the coordinate axes. The secular equation $\mathbf{A} - \lambda\mathbf{I} = 0$ has one zero and $N - 1$ positive roots

$$\lambda_j = 4 \sin^2 (j\pi/2N) \qquad (j = 1, 2, \cdots, N-1), \tag{7.7}$$

whence

$$P[2(N-1)\delta_{(1)}^2 \leq c^2] = P\left(\sum_1^{N-1} \lambda_j y_j^2 \leq c^2\right),$$

in which the y_j are mutually independent standardized normal variates. Note that the sth cumulant of $2(N-1)\delta_{(1)}^2$ is $(\sum_1^{N-1} \lambda_j^s)2^{s-1}(s-1)!$, by the additive property of cumulants. The sums of powers of the characteristic numbers of $\mathbf{A}$ required for the specification of the cumulants of $\delta_{(1)}^2$, may be expressed in terms of the minors of $\mathbf{A}$, using well-known results relating to symmetric functions of the roots of a polynomial equation or, alternatively, by direct summation of the finite trigonometric series $\sum_1^{N-1} \sin^{2s} (j\pi/2N)$, after expressing the powers of the trigonometric ratios in terms of trigonometric ratios of multiples of the angles by standard formulae.

Similar results apply to higher order successive differences, useful in eliminating the inflationary effect of suspected given polynomial trends on estimates of variability [30], [31]. The mean square kth order advancing difference $\delta_{(k)}^2$ is defined as

$$\begin{aligned} (N-k)\binom{2k}{k}\delta_{(k)}^2 &= \sum_{i=1}^{N-k} (\Delta^k x_i)^2 \\ &= \sum_{i=1}^{N-k}\left[\sum_{\alpha=0}^{k} (-)^\alpha \binom{k}{\alpha} x_{i+\alpha}\right]^2 \end{aligned} \qquad (k = 1, 2, \cdots, N-1). \tag{7.8}$$

The matrix of the quadratic form involved in the definition of $\delta_{(k)}^2$ is a continuant of order k in the sense that all the elements other than those in the leading

diagonal and in the secondary upper and lower k diagonals are zero. The matrix is of rank $N - k$ and

$$P\left((N-k)\binom{2k}{k}\delta^2_{(k)} \leqq c^2\right) = P\left(\sum_{j=1}^{N-k} \lambda_j y_j^2 \leqq c^2\right),$$

where the λ_j are the $N - k$ non-zero characteristic numbers of the matrix and the y_j are mutually independent standardized normal variates. The distribution of $\delta^2_{(2)}$ has been considered in some detail by Kamat [32].

(v) One additional application is worth noting. Consider a dynamic programming or multifactorial design set-up in which the optimal course of action is represented by the N-dimensional vector $\mathbf{x}^*$. Suppose that $\mathbf{x}^*$ is not known exactly due either to a penumbra of vagueness surrounding the model from which $\mathbf{x}^*$ is deduced (i.e. faulty or imperfect theory), or to the fact that $\mathbf{x}^*$ is predicated on past experience (i.e. limited sampling), or for some other reason. Denote the estimate of $\mathbf{x}^*$ by $\hat{\mathbf{x}}^*$, and let the expectation vector and variance-covariance matrix of $\hat{\mathbf{x}}^*$ be $\mathbf{x}^*$ and $\mathbf{V}(\hat{\mathbf{x}}^*)(\mathbf{V}(\hat{\mathbf{x}}^*)$ of full rank). A course of action $\mathbf{x}$ will be adopted aiming to approach $\mathbf{x}$ as closely as possible to the *assumed* ideal course of action $\hat{\mathbf{x}}^*$ (not to $\mathbf{x}^*$ which is unknown). Due to imperfect control of the action variables exact coincidence is not possible. Assume that the expectation vector and variance-covariance matrix of $\mathbf{x}$ are $\hat{\mathbf{x}}^*$ and $\mathbf{V}(\mathbf{x})$, respectively ($\mathbf{V}(\mathbf{x})$ of full rank). Then $\mathbf{d} = \mathbf{x} - \mathbf{x}^* = (\mathbf{x} - \hat{\mathbf{x}}^*) + (\hat{\mathbf{x}}^* - \mathbf{x}^*)$ has zero expectation vector and variance-covariance martix $\mathbf{V}(\hat{\mathbf{x}}^*) + \mathbf{V}(\mathbf{x})$, provided the two kinds of errors are uncorrelated. Let the loss function due to imperfect matching of $\mathbf{x}$ with $\mathbf{x}^*$ be the quadratic $\mathbf{d}'\mathbf{A}\mathbf{d}$, $|\mathbf{A}| > 0$, and assume further that $\mathbf{x}$, $\hat{\mathbf{x}}^*$ and therefore $\mathbf{d}$ have multivariate normal distributions. In view of the discussion in (iii), it is clear that the probability of the loss not exceeding a given upper bound c^2 is equal to $P(\sum_1^N \lambda_j y_j^2 \leqq c^2)$, in which the y_j have the usual significance. (In particular, the expected loss is $\sum_1^N \lambda_j$.) The reader is referred to Grad and Solomon [27] who discuss an analogous ballistics problem for which $N = 3$.

(vi) There is one interesting case for which the distribution of the weighted sum of squares may be expressed in exact form. If the number of components N is even, $N = 2m$, and the weights c_j coincides in pairs, say $c_j = c_{N-j}$, then

$$z = \sum_{j=1}^{N} c_j x_j^2 = \sum_{j=1}^{m} c_j y_j$$

where the y_j are independent χ^2, each with 2 degrees of freedom. The characteristic function of $\sum c_j y_j$ is $\Pi(1 - 2c_j it)^{-1}$. By partial fraction decomposition the latter may be expressed in the form $\sum A_j/(1 - 2c_j it)$, which is obviously directly invertible to $\sum (A_j/2c_j) \exp(-2z/c_j)$, the density function of z. It follows that the complement of the distribution function of z, $P(z > z_0)$, is likewise expressible as a linear combination of exponentials.

The above remarks may easily be extended to the situation where the weights are repeated in groups of four (instead of groups of two), groups of six, etc., i.e., to the situation where $z = \sum c_j x_j^2$ may be identified as a weighted sum of independent χ^2 variates, each with the same *even* number of degrees of freedom.

More generally still, the degrees of freedom of the components, though still even, need not be the same. A partial fraction representation of the characteristic function enables the density function to be expressed as a linear combination of Gamma density functions with degrees of freedom $2, 4, \cdots, p$, where p is the highest degree of freedom of the several components. It follows that the distribution function of the sum is a linear combination of Gamma distribution functions with degrees of freedom $2, 4, \cdots, p$.

We now obtain the recurrence relationship referred to at the beginning of this section. Note first that the intersection of the flat $x_N = x$ with the N-dimensional ellipsoid $\sum_1^N a_i x_i^2 \leqq t$ is itself an elliposid but of dimensionality $N - 1$ and with semi-axes of lengths $((t - a_N x^2)/a_i)^{\frac{1}{2}}$, $i = 1, 2, \cdots, N - 1$. The amount of probability within the ellipsoid intercepted by two parallel and adjoining flats $x_N = x$ and $x_N = x + dx$ is therefore

$$(2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}x^2)\, dx \cdot F_{N-1;b_1,\cdots,b_{N-1}}\left[\frac{t - a_N x^2}{\sum_1^{N-1} a_i}\right],$$

where $b_i = a_i/\sum_1^{N-1} a_j$, $0 < a_N < 1 (i = 1, 2, \cdots, N - 1)$. Consequently, the probability content of the ellipsiod is

$$\begin{aligned} &F_{N;a_1,\cdots a_N}(t) \\ (7.9) \qquad &= 2\int_0^{(t/a_N)^{\frac{1}{2}}} (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2} F_{N-1;b_1,\cdots,b_{N-1}}\left[\frac{t - a_N x^2}{1 - a_N}\right] dx \quad (N = 2, 3, \cdots), \end{aligned}$$

or, on setting $y = x(a_N/t)^{\frac{1}{2}}$,

$$\begin{aligned} &F_{N;a_1,\cdots,a_N}(t) \\ (7.10) \qquad &= 2\left(\frac{t}{a_N}\right)^{\frac{1}{2}} \int_0^1 (2\pi)^{-\frac{1}{2}} \exp\left(\frac{-ty^2}{2a_N}\right) \\ &\qquad \cdot F_{N-1;b_1,\cdots,b_{N-1}}\left[\frac{t(1 - y^2)}{1 - a_N}\right] dy \quad (N = 2, 3, \cdots).^{4a} \end{aligned}$$

We may note that for the particular case of $N - 1$ equal components,

$$(7.11) \qquad F_{N;\alpha,\cdots,\alpha,\beta}(t) = 2\left(\frac{t}{\beta}\right)^{\frac{1}{2}} \int_0^1 (2\pi)^{-\frac{1}{2}} \exp\left(\frac{-ty^2}{2\beta}\right) F_{N-1}\left[\frac{t(1 - y^2)}{1 - \beta}\right] dy,$$

in which $F_{N-1}(\cdot)$ denotes the distribution function of a χ^2 with $N - 1$ degrees of freedom, and $(N - 1)\alpha + \beta = 1$.

Finally, it will be convenient to record here an interesting relationship between the distribution of the weighted sum of squares of two independent standardized normal variables and that of the non-central χ^2 with two degrees of freedom. The relationship[5] in question is

[4a] (7.10) is, of course, just a convolution formula. It has been obtained here by a geometrical argument for consistency.

[5] I am indebted to one of the referees for having brought this useful result, for which an unpublished geometrical proof has been obtained by Dr. David C. Kleinecke, to my attention.

(7.12) $$F_{2;a_1,a_2}(t) = G_{2;\kappa}(u^2) - G_{2;u}(\kappa^2),$$

where

$$\kappa = \tfrac{1}{2}\,|\,(t/a_1)^{\frac{1}{2}} - (t/a_2)^{\frac{1}{2}}\,|, \qquad u = \tfrac{1}{2}[(t/a_1)^{\frac{1}{2}} + (t/a_2)^{\frac{1}{2}}],$$

and $G_{2;\kappa}(\cdot)$ is the distribution of $\chi^2_{2;\kappa}$, the non-central χ^2 with two degrees of freedom and non-centrality parameter κ, as defined in Section 3.

8. Probability content of a regular simplex. Denote the probability content of a regular N-dimensional simplex with sides of length a and centroid at the center of the spherical normal distribution with the same dimensionality by $K_N(a)$. To derive a formula for $K_N(a)$ it will be convenient to divide the simplex into $N + 1$ (non-regular) simplices, obtained by joining the centroid to the $N + 1$ vertices by straight lines.

Consider then one of these $N + 1$ derived simplices S. The probability content of this simplex may be obtained by first evaluating the amount of probability in a slab formed by two adjacent (N − 1)-flats parallel to the face opposite to that vertex of S which coincides with the centroid of the original simplex. Let x and $x + dx$ denote the distances of these flats from the latter vertex. The intersection of the first flat with S is a *regular* simplex of dimensionality $N - 1$ and with its edges of length y, where $y/a = x/d$, d denoting the distance of the centroid of the original simplex from one of its faces. Furthermore, the density at a point on the same flat distant ξ from the centroid of this regular simplex is

$$(2\pi)^{-\frac{1}{2}N} \exp(-\tfrac{1}{2}r^2) = (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}x^2).(2\pi)^{-\frac{1}{2}(N-1)} \exp(-\tfrac{1}{2}\xi^2),$$

where $r^2 = x^2 + \xi^2$ is the square of the distance of the centroid of the original simplex from the point in question. It follows that the probability content of the slab is $(2\pi)^{-\frac{1}{2}} \exp(-\frac{1}{2}x^2)\, dx K_{N-1}(ax/d)$, and the probability content of the original simplex is

(8.1) $$K_N(a) = (N + 1) \int_0^d (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}x^2) K_{N-1}\left(\frac{ax}{d}\right) dx.$$

It is easily shown that

(8.2) $$d = a/(2N(N + 1))^{\frac{1}{2}}.$$

Substituting for d in (8.1), the desired integral recurrence relationship is obtained:

(8.3) $$K_N(a) = (N + 1) \int_0^{a/(2N(N+1))^{\frac{1}{2}}} (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}x^2) K_{N-1}[x(2N(N + 1))^{\frac{1}{2}}]\, dx \quad (N = 1, 2, \cdots),$$

or, equivalently [34][6],

[6] Godwin actually uses functions $G_m(\cdot)$ which are related to the K-functions by $G_m(x/\sqrt{2}) = (2\pi)^{m/2}(m + 1)^{-\frac{1}{2}}K_m(x)$. The K-function, from a theoretical point of view, seems to be more convenient and natural than the G-function, since it is (unlike the latter) a distribution function, in the usual statistical sense.

$$K_N(a) = \tfrac{1}{2}\left(\frac{N+1}{N\pi}\right)^{\frac{1}{2}} \int_0^a \cdot \exp\left(-\frac{u^2}{4N(N+1)}\right) K_{N-1}(u)\, du \quad (N = 1, 2, \cdots), \tag{8.4}$$

with $K_0(a) = 1$.

We shall give now one application involving a knowledge of $K_N(a)$. This relates to the problem of determining the distribution of the sum (or mean) of N independent observations from a half-normal distribution, or equivalently the sum (or mean) of the absolute values of observations from a normal population. The first four moments of this distribution have been obtained by Kamat [33] for $N \leq 3$, but the actual distribution for general N does not appear to have been obtained previously.

The density function for each observation is $(2/\pi)^{\frac{1}{2}} \exp(-x^2/2)(0 \leq x < \infty)$, and the joint density function of N independent observations is

$$\left(\frac{2}{\pi}\right)^{\frac{1}{2}N} \exp\left(-\tfrac{1}{2}\sum_{i=1}^{N} x_i^2\right) \quad (x_i \geq 0, \quad i = 1, 2, \cdots, N). \tag{8.5}$$

The determination of the density function of $u = \sum_1^N x_i$ thus reduces to the determination of the probability intercepted by the $(N-1)$-flats $\sum_1^N x_i = u$ and $\sum_1^N x_i = u + du$ in the positive orthant. To obtain this, note that $\sum_1^N x_i = u$, $x_i \geq 0 (i = 1, 2, \cdots, N)$ defines a regular $(N-1)$-dimensional simplex with edges of length $u\sqrt{2}$. The distance of the flat $\sum_1^N x_i = u$ from the origin (i.e., the distance of the latter point from the centroid of the simplex) is $u/\sqrt{N}$. Further, the density at any point within the simplex distant η from the centroid may be expressed in the form

$$\left(\frac{2}{\pi}\right)^{\frac{1}{2}N} \exp\left(-\tfrac{1}{2}\sum_1^N x_i^2\right) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}N} \exp\left(-\tfrac{1}{2}\left[\frac{u^2}{N} + \eta^2\right]\right)$$

$$= \frac{1}{(2\pi)^{\frac{1}{2}(N-1)}} \exp(-\tfrac{1}{2}\eta^2)\, \frac{2^N}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2N}\right).$$

Consequently, the probability content of the element $u \leq \sum x_i \leq u + du$ is

$$h_N(u)\, du = \frac{2^N}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2N}\right) \frac{du}{\sqrt{N}} K_{N-1}(u\sqrt{2}), \tag{8.6}$$

after integration with respect to η over the simplex, where $h_N(u)$ denotes the p.d.f. of u. Observe that equation (8.6) reveals at the same time the intimate tie-up between the K_{N-1}-function and the N-fold convolution of the half-normal distribution. This tie-up was first demonstrated by Godwin [35], using an entirely different argument, in showing the equivalence of two expressions for the p.d.f. of the mean absolute deviation in normal samples, obtained respectively in [36] and [37].

We conclude with two further important applications of the K-function. The first relates (as already indicated) to the distribution of the mean absolute deviation in samples of size n from a normal population with zero mean and unit variance. The p.d.f., $p_n(t)$, of the latter variable is given by

$$p_n(t) = \left(\frac{n}{2}\right)^{\frac{1}{2}} \pi^{-\frac{1}{2}} \sum_{k=1}^{n-1} \binom{n}{k} [k(n-k)]^{-\frac{1}{2}} \exp\left[-\frac{n^2t^2}{8k(n-k)}\right] \cdot K_{k-1}\left(\frac{nt}{\sqrt{2}}\right) K_{n-k-1}\left(\frac{nt}{\sqrt{2}}\right), \tag{8.7}$$

(Godwin [36]). Tables of $\int_0^x p_n(t)\,dt$ for $n \leq 10$ are available in [38], while percentage points of the distribution are given in [38] and [39] (Table 21, p. 165).

The second application relates to the distribution of the deviation of the largest obser ation from the mean in a sample of n independent observations from a norm. l population with zero mean and unit variance. The distribution function, $Q_n(t)$, of the latter variable is given by

$$Q_n(t) = K_{n-1}(nt\sqrt{2}) \tag{8.8}$$

(Nair [40]). Tables of $Q_n(t)$ are given in [40], while percentage points of the distribution are available in [40] and [39] (Table 25, p. 172).

It should be noted that the K-functions also find applications in connection with the distributions of a class of linear functions of normal order statistics [40][7].

REFERENCES

[1] P. C. Mahalanobis, R. C. Bose and S. N. Roy, "Normalisation of statistical variates and the use of rectangular coordinates in the theory of sampling distributions," *Sankhyā*, Vol. 3 (1937), pp. 1–40.

[2] James Clerk Maxwell, "Illustrations of the dynamical theory of gases—Part 1. On the motions and collisions of perfectly elastic bodies," *Taylor's Philosophical Magazine*, Vol. 19 (1860), pp. 19–32.

[3] M. G. Kendall, *The Advanced Theory of Statistics*, Vol. 1, Charles Griffin and Co., London, 1946.

[4] P. B. Patnaik, "The non-central χ^2 and F-distributions and their application," *Biometrika*, Vol. 25 (1949), pp. 202–232.

[5] R. A. Fisher, "The general sampling distribution of the multiple correlation coefficient," *Proc. Roy. Soc.*, Series A, Vol. 121 (1928), pp. 654–673.

[6] F. Garwood, Unpublished Ph.D. Thesis, University of London, 1934.

[7] Evelyn Fix, *Tables of the Non-Central Chi-square Distribution, Vol. 1*. Univ. Calif. Publ. Stat., 1949.

[8] S. H. Abdel-Aty, "Approximate formulae for the percentage points and the probability integral of the non-central χ^2 distribution," *Biometrika*, Vol. 41 (1954), pp. 538–540.

[9] John W. Tukey, "Approximations to the upper 5% points of Fisher's B distribution and non-central χ^2," *Biometrika*, Vol. 44 (1957), pp. 528–531.

[10] J. I. Marcum, *Tables of Q-Functions*. Rand Report No. RM-339, Santa Monica, Calif., 1950.

[7] Still further (new) applications of the K-function will be shown elsewhere.

[11] H. H. Germond, *The Circular Coverage Function.* Rand Report No. RM-330, Santa Monica, Calif., 1950.
[12] *Offset Circle Probabilities.* Rand Report No. R-234, Santa Monica, Calif., 1952.
[13] *Tables of Circular Normal Probabilities.* Operations Analysis Group, Dynamics Section, Report No. 02-949-106, Bell Aircraft Corporation, 1956.
[14] British Association, *Mathematical Tables, Vol. 1.* Cambridge University Press, 1931.
[15] Cecil C. Craig, "Note on the distribution of the non-central t with an application," *Ann. Math. Stat.*, Vol. 12 (1941), pp. 224–228.
[16] J. Neyman and B. Tokarska, "Errors of the second kind in testing Student's hypothesis," *J. Amer. Stat. Soc.*, Vol. 31 (1936), pp. 318–326.
[17] N. L. Johnson and B. L. Welch, "Applications of the non-central t-distribution," *Biometrika*, Vol. 31 (1940), pp. 362–389.
[18] G. J. Resnikoff and G. J. Lieberman, *Tables of the Non-Central t-Distribution.* Stanford Univ. Press, 1957.
[19] D. M. Y. Sommerville, *An Introduction to the Geometry of N Dimensions.* Methuen and Co. Ltd., London, 1922.
[20] A. Bhattacharya, "A note on the distribution of the sum of chi-squares," *Sankhyā*, Vol. 7 (1945), pp. 27–28.
[21] Herbert Robbins, "The distribution of a definite quadratic form," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 266–270.
[22] Herbert Robbins and E. J. G. Pitman, "Application of the method of mixtures to quadratic forms in normal variables," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 552–560.
[23] Harold Hotelling, "Some new methods for distribution of quadratic forms," *Ann. Math. Stat.*, Vol. 19 (1948), Abstract.
[24] John Gurland, "Distribution of quadratic forms and some applications," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 416–427.
[25] ——, "Distribution of definite and of indefinite quadratic forms," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 122–127.
[26] James Pachares, "Note on the distribution of a definite quadratic form," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 728–731.
[27] Arthur Grad and Herbert Solomon, "Distribution of quadratic forms and some applications," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 464–477.
[28] J. von Neumann, H. R. Bellinson, R. M. Kent and B. I. Hart, "The mean square successive difference," *Ann. Math. Stat.*, Vol. 12 (1941), pp. 153–162.
[29] J. von Neumann, "Distribution of the ratio of the mean square successive difference to the variance," *Ann. Math. Stat.*, Vol. 12 (1941), pp. 367–395.
[30] G. Tintner, *The Variate-Difference Method.* Principia Press, Bloomington, Indiana, 1940.
[31] Anthony P. Morse and Frank E. Grubbs, "The estimation of dispersion from differences," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 194–214.
[32] A. R. Kamat, "Distribution theory of two estimates for standard deviation based on second variate differences," *Biometrika*, Vol. 41 (1954), pp. 1–11.
[33] ——, "Incomplete and absolute moments of the multivariate normal distribution," *Biometrika*, Vol. 40 (1953), pp. 20–34.
[34] H. J. Godwin, "A further note on the mean deviation," *Biometrika*, Vol. 35 (1948), pp. 304–309.
[35] ——, "A note on Kac's deviation of the distribution of the mean deviation," *Ann. Math. Stat.*, Vol. 20 (1949), p. 127.
[36] ——, "On the distribution of the estimate of the mean deviation obtained from samples from a normal population," *Biometrika*, Vol. 33 (1945), pp. 254–256.
[37] M. Kac, "On the characteristic functions of the distributions of various deviations

in samples from a normal population," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 257–261.

[38] H. J. Godwin and H. O. Hartley, "Tables of the probability integral of the mean deviation in normal samples," *Biometrika*, Vol. 33 (1945), pp. 259–265.

[39] E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians, Vol. 1*. Cambridge Univ. Press, 1954.

[40] K. R. Nair, "The distribution of the extreme deviate from the sample mean and its studentised form," *Biometrika*, Vol. 35 (1948), pp. 118–144.

GRAPHS FOR BIVARIATE NORMAL PROBABILITIES

By Marvin Zelen and Norman C. Severo[1]

National Bureau of Standards

1. Introduction and summary. Recently there has been much activity dealing with the tabulation of the bivariate normal probability integral. D. B. Owen [3], [4] has summarized many of the properties of the bivariate normal distribution function and tabulated an auxiliary function which enables one to calculate the bivariate normal probability integral. In addition, the National Bureau of Standards [1] has compiled extensive tables of the bivariate normal integral drawn from the works of K. Pearson, Evelyn Fix and J. Neyman, and H. H. Germond. In this same volume, D. B. Owen has contributed an extensive section on applications.

It is the purpose of this paper to present three charts, which will enable one to easily compute the bivariate normal integral to a maximum error of 10^{-2}. This should be sufficient for most practical applications. Owen and Wiesen [5] have also presented charts with a similar objective; however, as pointed out below, we believe the charts presented here lend themselves more easily to visual interpolation. Actually the motivation for these charts came from the Owen and Wiesen work.

2. Notation and formulas. We present here notation and useful formulas relating to the bivariate normal integral. Let X and Y be random variables following a bivariate normal distribution with zero means, unit variances, and correlation coefficient ρ. Then

$$Pr\{X \geqq h, Y \geqq k\} = L(h, k; \rho) = \int_h^\infty dx \int_k^\infty g(x, y; \rho)\, dy, \tag{1}$$

where

$$g(x, y; \rho) = [2\pi\sqrt{1-\rho^2}]^{-1} \exp - \tfrac{1}{2}[(x^2 - 2\rho xy + y^2)/(1-\rho^2)]$$

is the bivariate normal probability density function.

Useful relations for $L(h, k; \rho)$ are set out below:

$$L(h, k; \rho) = L(k, h; \rho), \tag{2}$$

$$L(-h, -k; \rho) = \int_{-\infty}^h dx \int_{-\infty}^k g(x, y; \rho)\, dy, \tag{3}$$

$$L(-h, k; -\rho) = \int_{-\infty}^h dx \int_k^\infty g(x, y; \rho)\, dy, \tag{4}$$

Received September 30, 1959.

[1] Now at the University of Buffalo.

$$(5)\qquad L(h, -k; -\rho) = \int_h^\infty dx \int_{-\infty}^k g(x, y; \rho)\, dy,$$

$$(6)\qquad 2[L(h, k; \rho) + L(h, k; -\rho) + P(h) - Q(k)] = \int_{-h}^h dx \int_{-k}^k g(x, y; \rho)\, dy,$$

$$(7)\qquad L(-h, k; \rho) + L(h, k; -\rho) = Q(k),$$

$$(8)\qquad L(-h, -k; \rho) - L(h, k; \rho) = P(k) - Q(h),$$

where

$$P(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^x e^{-t^2/2}\, dt = 1 - Q(x).$$

Special values of $L(h, k; \rho)$ are

$$(9)\qquad L(h, k; 0) = Q(h)Q(k),$$

$$(10)\qquad L(h, k; -1) = 0 \qquad \text{if } h + k \geqq 0,$$

$$(11)\qquad L(h, k; -1) = P(h) - Q(k) \qquad \text{if } h + k \leqq 0,$$

$$(12)\qquad L(h, k; 1) = Q(h) \qquad \text{if } k \leqq h,$$

$$(13)\qquad L(h, k; 1) = Q(k) \qquad \text{if } k \geqq h,$$

$$(14)\qquad L(0, 0; \rho) = \tfrac{1}{4} + ((\arcsin \rho)/2\pi).$$

3. Discussion of Charts. Owen [4] has shown that

$$(15)\qquad L(h, k; \rho) = L\left(h, 0; \frac{(\rho h - k) \operatorname{sgn} h}{(h^2 - 2\rho hk + k^2)^{\frac{1}{2}}}\right) + L\left(k, 0; \frac{(\rho k - h) \operatorname{sgn} k}{(h^2 - 2\rho hk + k^2)^{\frac{1}{2}}}\right)$$
$$- \begin{cases} 0 & \text{if } \quad hk > 0 \quad \text{or} \quad hk = 0 \quad \text{and} \quad h + k \geqq 0 \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

This makes it possible to evaluate $L(h, k; \rho)$ as a function of $L(h, 0; \rho)$ which only depends on two parameters. Figures 1, 2, and 3 are plots of h versus ρ with constant contour lines such that $L(h, 0; \rho) = 0.01(.01).10(.02).50$.

Owen and Wiesen [5] have given charts plotting $L(h, 0; \rho)$ versus h with constant contours for ρ. The advantages of the charts presented here are that (i) the contour lines more fully cover the available graph space making interpolation easier and more accurate, and (ii) the Owen and Wiesen charts require visual interpolation on the ρ contour lines which could easily lead to errors larger than $\pm.01$ in reading L. On the other hand, figures 1, 2, and 3 require only visual interpolation on L between successive contour lines differing by .01 or .02. Hence interpolation errors are at most of the order $\pm.01$ throughout.

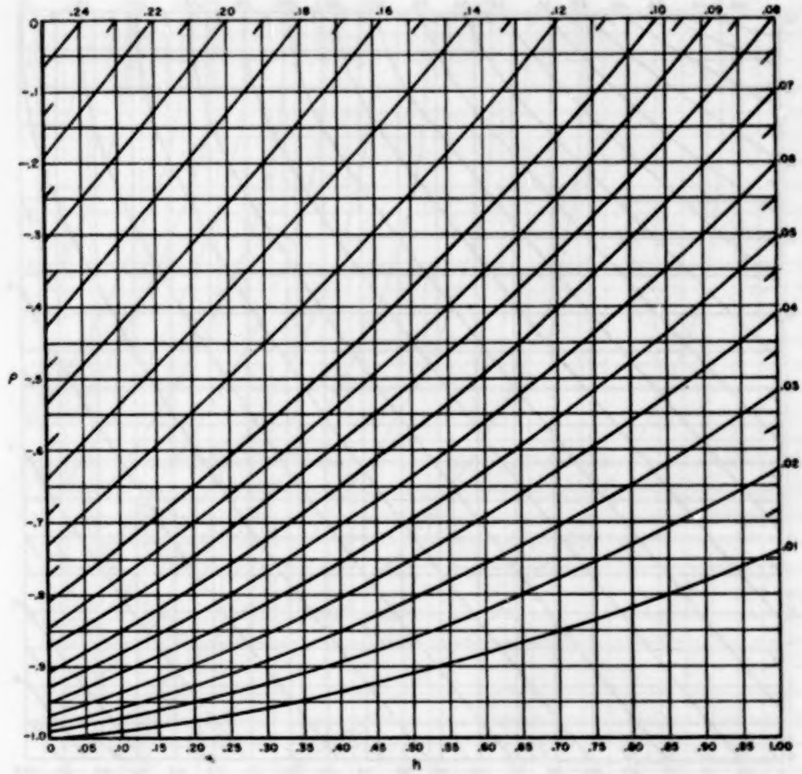


FIG. 1. $L(h, 0; \rho)$ for $0 \leqq h \leqq 1$ and $-1 \leqq \rho \leqq 0$. Values for $h < 0$ can be obtained using

$$L(h, 0; -\rho) = \tfrac{1}{2} - L(-h, 0; \rho)$$

4. Applications of the charts.

EXAMPLE 1. To find $L(0.5, 0.4; 0.8)$. Using (15), we have

$$(h^2 - 2\rho hk + k^2)^{\frac{1}{2}} = 0.3$$

$$L(0.5, 0.4; 0.8) = L(0.5, 0; 0) + L(0.4, 0; -.6) = 0.15 + 0.08 = 0.23.$$

The correct answer to 3D from [1] is $L(0.5, 0.4; 0.8) = 0.233$.

EXAMPLE 2. Let X and Y follow a bivariate normal distribution with means and variances $m_x = 3$, $m_y = 2$, $\sigma_x^2 = 16$, $\sigma_y^2 = 4$ and correlation $\rho = -0.125$. To find the value of $Pr\{X \geqq 2, Y \geqq 4\}$. Since

$$Pr\{X \geqq h, Y \geqq k\} = L[(h - m_x)/\sigma_x, (k - m_y)/\sigma_y; \rho],$$

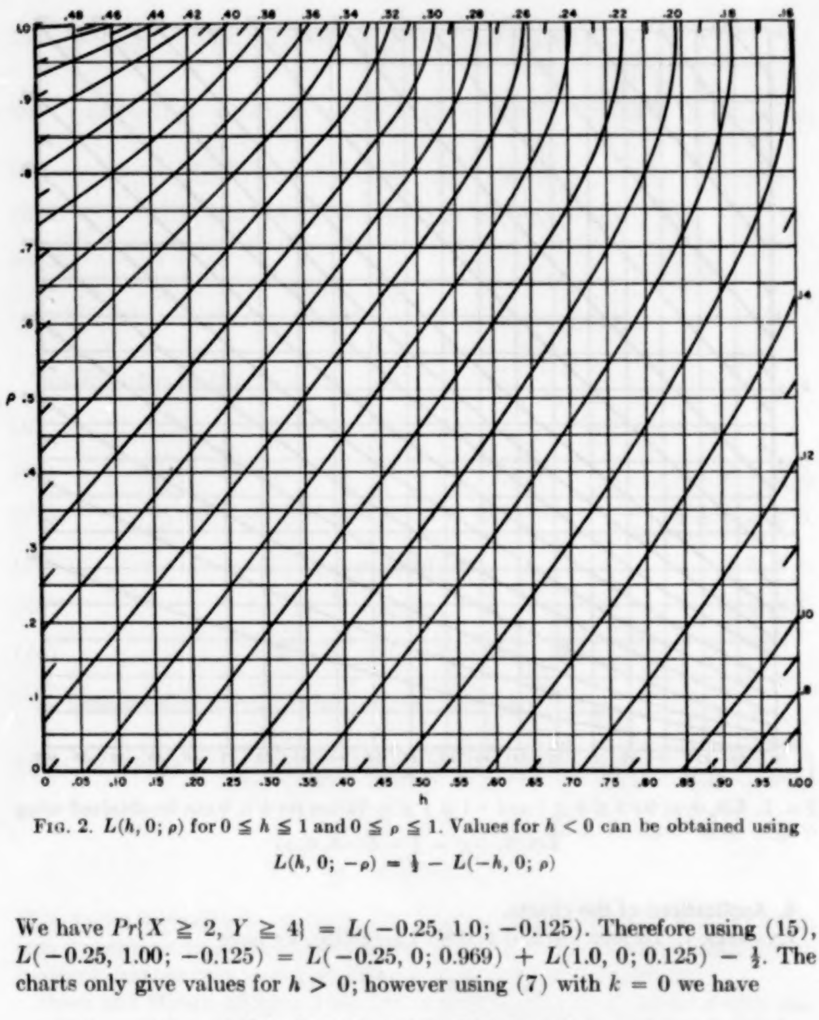


FIG. 2. $L(h, 0; \rho)$ for $0 \leqq h \leqq 1$ and $0 \leqq \rho \leqq 1$. Values for $h < 0$ can be obtained using $L(h, 0; -\rho) = \frac{1}{2} - L(-h, 0; \rho)$

We have $Pr\{X \geqq 2, Y \geqq 4\} = L(-0.25, 1.0; -0.125)$. Therefore using (15), $L(-0.25, 1.00; -0.125) = L(-0.25, 0; 0.969) + L(1.0, 0; 0.125) - \frac{1}{2}$. The charts only give values for $h > 0$; however using (7) with $k = 0$ we have

$$L(-h, 0, \rho) = \tfrac{1}{2} - L(h, 0; -\rho).$$

Hence $L(0.25, 0; .969) = \frac{1}{2} - L(0.25, 0; -.969)$ and thus

$$L(-0.25, 1.0; -.125) = -\mathrm{L}(0.25, 0; -.969) + L(1.0, 0; 0.125)$$

$$= -0.01 + 0.09 = 0.08.$$

The correct answer to 3D from [1] is $L(-0.25, 1.0; -.125) - 0.080$.

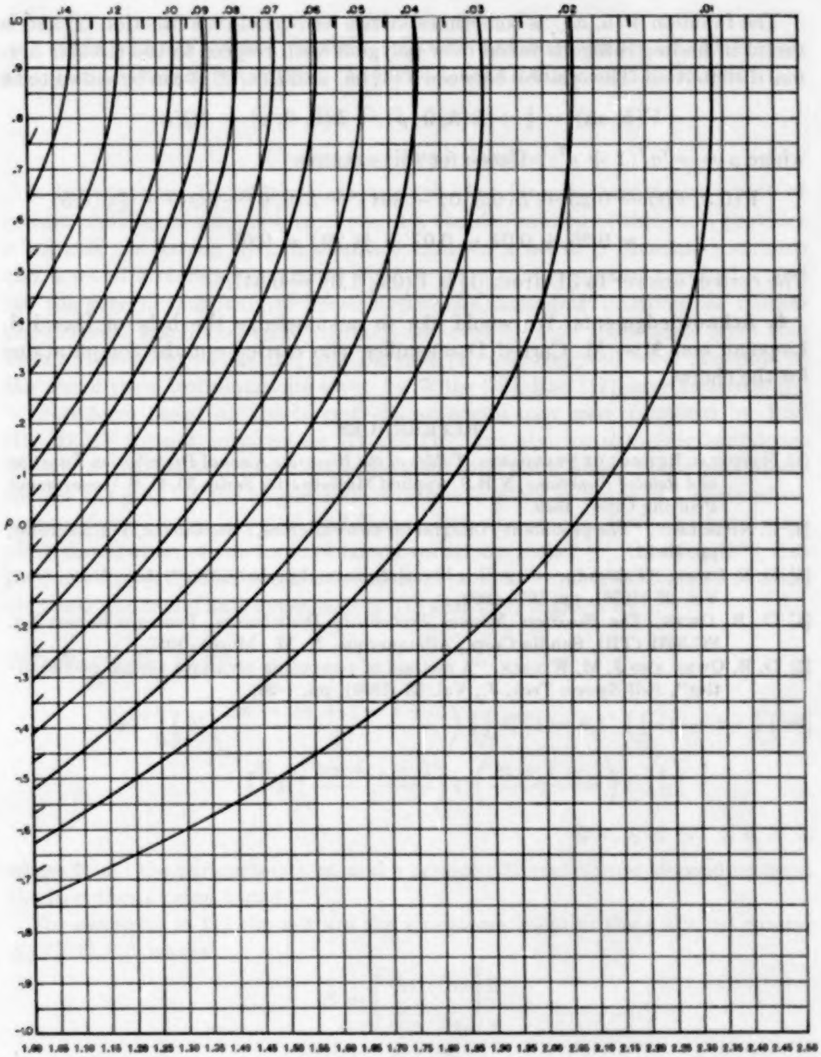


FIG. 3. $L(h, 0; \rho)$ for $h \geqq 1$ and $-1 \leqq \rho \leqq 1$. Values for $h < 0$ can be obtained using

$$L(h, 0; -\rho) = \tfrac{1}{2} - L(-h, 0; \rho)$$

EXAMPLE 3. To find the value of

$$V(h, ah) = (2\pi)^{-1} \int_0^h dx \int_0^{ax} e^{-\frac{1}{2}(x^2+y^2)}\, dy$$

when $a = 2$ and $h = 0.5$.

The function $V(h, ah)$ is sometimes known as Nicholson's function [2] and is useful in finding integrals taken over polygons with respect to the circular normal distribution. The relation between $V(h, ah)$ and $L(h, 0; \rho)$ can be shown to be

$$V(h, ah) = \tfrac{1}{4} + L(h, 0; \rho) - L(0, 0; \rho) - \tfrac{1}{2}Q(h)$$

where $\rho = -[a/(1 + a^2)]^{\frac{1}{2}}$. Hence for this example

$$\begin{aligned} V(0.5, 1.0) &= 0.25 + L(0.5, 0; -.894) - L(0, 0; -.894) - \tfrac{1}{2}Q(0.5) \\ &= 0.25 + 0.01 - 0.07 - \tfrac{1}{2}(.29) = 0.05 \end{aligned}$$

The correct answer to 3D from [1] is $V(0.5, 1.0) = 0.047$.

5. Acknowledgments. We would like to acknowledge the help of David S. Liepman and Miss M. Carroll Dannemiller who carried out the computations for the charts.

REFERENCES

[1] National Bureau of Standards, *Tables of the Bivariate Normal Distribution Function and Related Functions*, N.B.S. Applied Mathematics Series 50, U. S. Government Printing Office, 1959.

[2] C. Nicholson, "The probability integral for two variables," *Biometrika*, Vol. 33 (1943), pp. 59–72.

[3] D. B. Owen, "Tables for computing bivariate normal probabilities", *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1075–1090.

[4] D. B. Owen, *The Bivariate Normal Probability Distribution*, Research Report No. SC-3831 (TR), Sandia Corp., Albuquerque, N. M., March, 1957.

[5] D. B. Owen and J. M. Wiesen, "A method of computing bivariate normal probabilities", *Bell System Tech. J.*, Vol. 38 (1959), pp. 1–20.

CHARTS OF SOME UPPER PERCENTAGE POINTS OF THE DISTRIBUTION OF THE LARGEST CHARACTERISTIC ROOT[1]

By D. L. Heck[2]

University of North Carolina

1. Introduction. In multivariate analysis, the largest characteristic root of certain matrices of sample quantities, or a simple function of this root, provides a statistic for testing (i) independence between a set of p correlated variates and a set of q correlated variates in a $(p + q)$-variate normal population and (ii) the general multivariate linear hypothesis, assuming multivariate normal populations. Likelihood ratio methods for dealing with these tests have also been advanced by Wilks [29] and Bartlett [4], and comprehensive accounts of the use of these techniques are given by Wilks [30], Rao [17], and Anderson [2].

Procedures based on the largest characteristic root were proposed by Roy [19, 20, 23] for not only testing (i) and (ii), but also for obtaining confidence bounds on parametric functions associated with both cases. These procedures require the c.d.f. of the largest root, which was given in terms of a chain of recursion formulae by Roy [19] and Nanda [12] who started from the joint sampling distribution of the roots obtained earlier by Fisher [6], Girshick [10], Hsu [11], and Roy [18]. This distribution of θ_i $(i = 1, 2, \cdots, s)$, the s non-zero roots obtained under the null hypothesis in (i) and (ii), is given by

$$p(\theta_1, \theta_2, \cdots, \theta_s) \prod_{i=1}^{s} d\theta_i$$

$$= \frac{\pi^{s/2} \prod_{i=1}^{s} \Gamma\left(\frac{2m + 2n + s + i + 2}{2}\right) \prod_{i=1}^{s} \theta_i^m (1 - \theta_i)^n \prod_{i>j}^{s} (\theta_i - \theta_j) \prod_{i=1}^{s} d\theta_i}{\prod_{i=1}^{s} \Gamma\left(\frac{2m + i + 1}{2}\right) \Gamma\left(\frac{2n + i + 1}{2}\right) \Gamma\left(\frac{i}{2}\right)}$$

$$0 < \theta_1 \leqq \cdots \leqq \theta_s < 1,$$

where θ_i and the parameters s, m, and n assume different values, depending upon the hypothesis being tested.

For example, in (i) the θ_i's are the s non-zero roots of the $(p \times p)$ matrix $S_{12}S_{22}^{-1}S_{12}'S_{11}^{-1}$, where

$$S = \begin{array}{cc} \left[\begin{array}{cc} S_{11} & S_{12} \\ S_{12}' & S_{22} \end{array}\right] & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & \;\; q \end{array} & \end{array}$$

Received April 29, 1958; revised May 25, 1959.

[1] This paper was prepared with the support of the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command. Reproduction in whole or part is permitted for any purpose of the United States Government.

[2] Now with International Business Machines Corporation, Endicott, New York.

is the sample covariance matrix based on $N - 1$ degrees of freedom. The parameters are given by $s = \min(p, q)$, $m = (|p - q| - 1)/2$, and

$$n = (N - p - q - 2)/2.$$

In (ii) and the special case of testing for equality of the $(p \times 1)$ mean vectors of k groups in one-way classification multivariate analysis of variance, $\theta_i = c_i/(1 + c_i)$, where the c_i's are the s non-zero roots of the $(p \times p)$ matrix BW^{-1}. The elements of the $(p \times p)$ matrix B consist of the "between groups" corrected sums of squares and crossproducts of the p variates, while W is the corresponding matrix for "within groups". With n_j observations on the p variates in each of the k groups, $s = \min(k - 1, p)$, $m = (|k - p - 1| - 1)/2$, and $n = (\sum_{j=1}^{k} n_j - k - p - 1)/2$.

The tests are carried out by computing θ_s and then comparing this statistic with the appropriate 100α percentage point of the distribution of the largest root. Further applications of these percentage points for testing purposes have been given by Bargmann [3], Chaudhuri [5], Foster and Rees [7], Pillai [15], Roy [20], [23], Roy and Bargmann [25], and Roy and Roy [27]. In addition, the use of the points in setting up multivariate confidence bounds is discussed in [3], [21], [22], [23], [24], [25], [26], [27].

The purpose of this paper is to present, in chart form, some upper percentage points of the distribution of the largest root for a wider range of the parameters than has heretofore been considered. One of the most extensive tabulations to date is that by Pillai [15], giving the upper 1% and 5% points and covering the range $s = 2(1)5$, $m = 0(1)4$, $n = 5(5)40(20)100(30)160, 200, 300, 500, 1000$. Also, the upper 1% and 5% points for these same values of m and n have been obtained by Pillai and Bantegui [16] for $s = 6$. Other tables include Nanda's [13], the upper 1% and 5% points for $s = 2$, $m = 0(\frac{1}{2})2$, $n = \frac{1}{2}(\frac{1}{2})10$; Chaudhuri's [5], the upper 1% and 5% points for $s = 2$, $m = n = 2\frac{1}{2}(\frac{1}{2})5(1)11$, for $s = 3$, $m = n = 2\frac{1}{2}(\frac{1}{2})5(1)8$, and for $s = 3$, $m = 0(\frac{1}{2})2$, $n = \frac{1}{2}(\frac{1}{2})2$; Foster and Rees' [7], the upper 1%, 5%, 10%, 15%, and 20% points for $s = 2$, $m = -\frac{1}{2}$, $0(1)9$, $n = 1(1)19(5)49, 59, 79$; and Foster's [8, 9], the upper 1%, 5%, 10%, 15%, and 20% points for $s = 3, 4$, $m = -\frac{1}{2}(\frac{1}{2})3$, $n = 0(1)95$. From Table 4.1 and the charts in Section 3, the upper 1%, 2.5%, and 5% points may be obtained for $s = 2(1)5$, $m = -\frac{1}{2}, 0(1)10$, $n \geq 5$.

2. Computation of the percentage points. The charts in Section 3 were prepared from percentage points which were computed using two types of approximations to the c.d.f. of the largest characteristic root. The first type of approximation, obtained for $s = 2, 3, 4, 5$ by Pillai [14] was used to compute, in general, the points for integral $n \leq 100$. For large values of n, generally $n > 100$ asymptotic approximations based on Pillai's formulae were used which were obtained by Whittlesey [28].

To compute the percentage points from Pillai's approximations, denoted by $p_s(x, m, n)$, the value of $p_s(x, m, n)$ for a particular combination (s, m, n) was

first calculated at the 100 values of x from .01 to 1.0 at intervals of .01. On the resulting ordinates, a method of inverse interpolation was used to obtain the upper 1%, 2.5%, and 5% points, i.e. x_α such that

$$p_s(x_\alpha, m, n) = 1 - \alpha \qquad (\alpha = .01, .025, .05).$$

The overall computational procedure for each value of s was as follows: For a fixed integral m and an initial (small) n, the percentage points were computed; n was then stepped up by unit increments, with the percentage points being computed for each value of n, until the desired set of values of n was covered. Then the expression was modified for the next integral value of m, and the percentage points for this value of m were computed for all desired n. This procedure was continued to $m = 10$, which is a fairly large value for practical purposes.

As a check on the accuracy of these percentage points, a number of the points were substituted in the expression for the exact c.d.f., and the largest error which occurred was found to be less than two units in the fourth decimal.

Whittlesey's asymptotic approximations (for integral values of m) may be obtained from Pillai's approximations by using Stirling's approximation and the substitution

$$z = -(m + 2n + s + 1) \log (1 - x), \tag{2.1}$$

and then letting n become large. From the resulting expressions, denoted by $w_s(z, m)$, inverse interpolation was used to obtain $z_\alpha(s, m)$(or z_α) such that for fixed s and m,

$$w_s(z_\alpha, m) = 1 - \alpha \qquad (\alpha = .01, .025, .05).$$

From these "asymptotic" $z_\alpha(s, m)$ values, given in Table 4.1, the percentage points $x_\alpha(s, m, n)$ were obtained by inverting (2.1).

A group of the percentage points obtained from Whittlesey's approximations was checked by substitution in the expression for the exact c.d.f., and of those points used in the final tabulation, the error for the most unfavorable combination of s and m ($s = 5, m = 10$) was found to be five units in the fourth decimal. This error, which is primarily an error of asymptotic approximation, is considerably smaller for smaller values of s and m, and, because of the asymptotic nature of the approximation, decreases in all cases, for increasing n.

Computation of the percentage points and the $z_\alpha(s, m)$ values was carried out on the IBM 650, with the programs coded in The Bell Interpretive System [31]. The program of the exact c.d.f. of the largest root, which was used for checking purposes, was coded in DOPSIR [1] (for $s = 2(1)6$, $m = 0(1)10$, and integral $n \geqq 0$), and is available at the North Carolina State College IBM Laboratory, The Institute of Statistics, Raleigh, North Carolina. The computation of the points for $m = -\frac{1}{2}$ was done subsequent to the computation for integral valued m, and Pillai's and Whittlesey's approximations were again used, after appropriate modifications were made.

CHART I

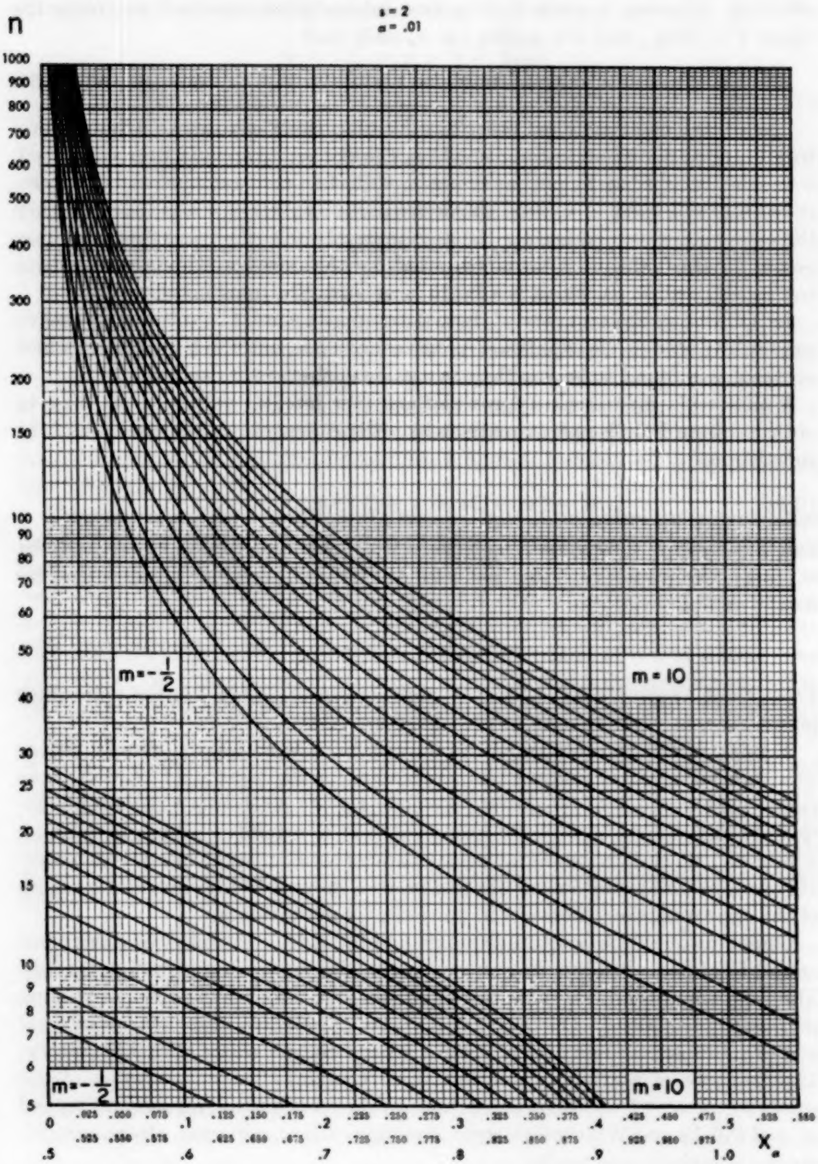


CHART II

s = 2

α = .025

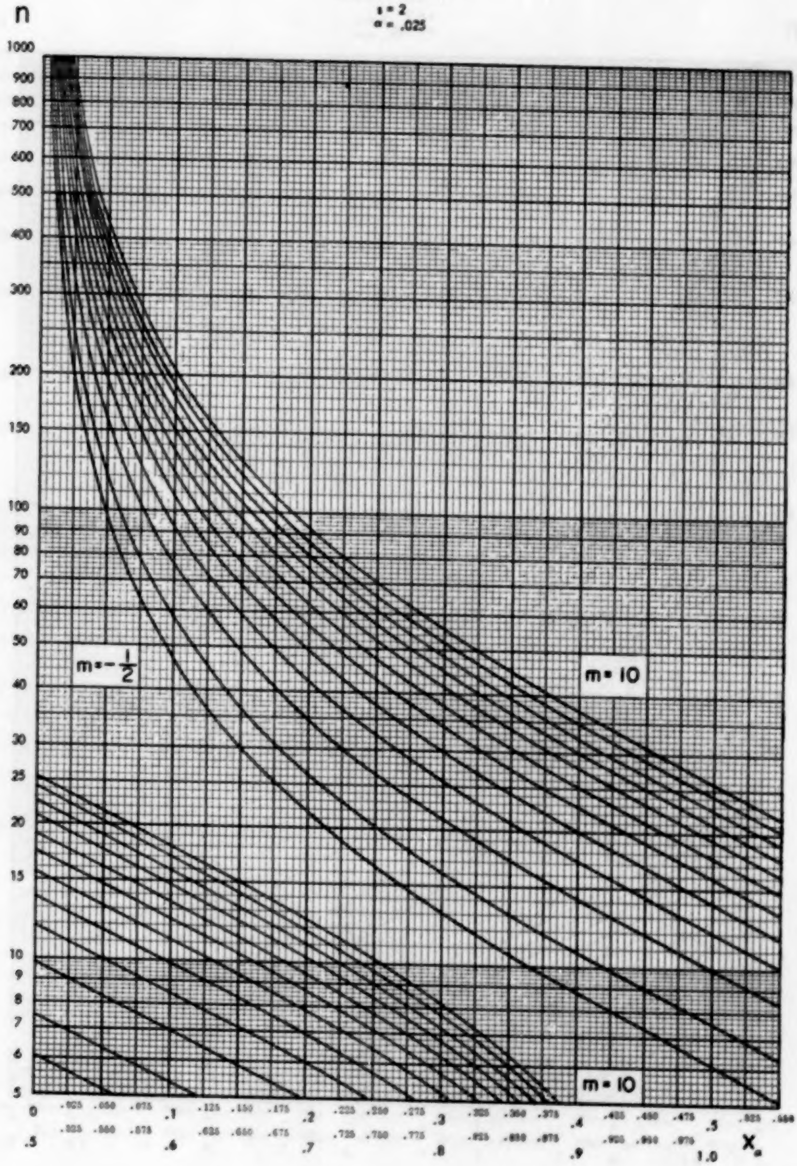


CHART III

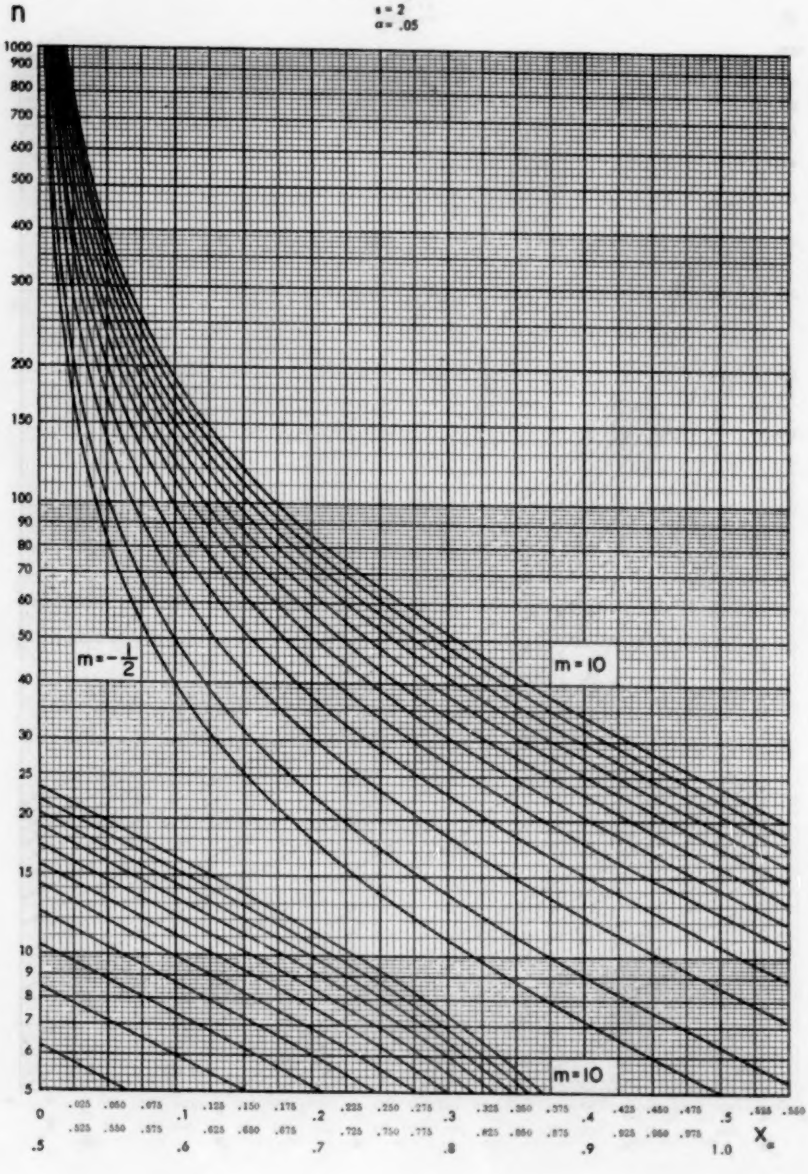


CHART IV

$s = 3$
$\alpha = .01$

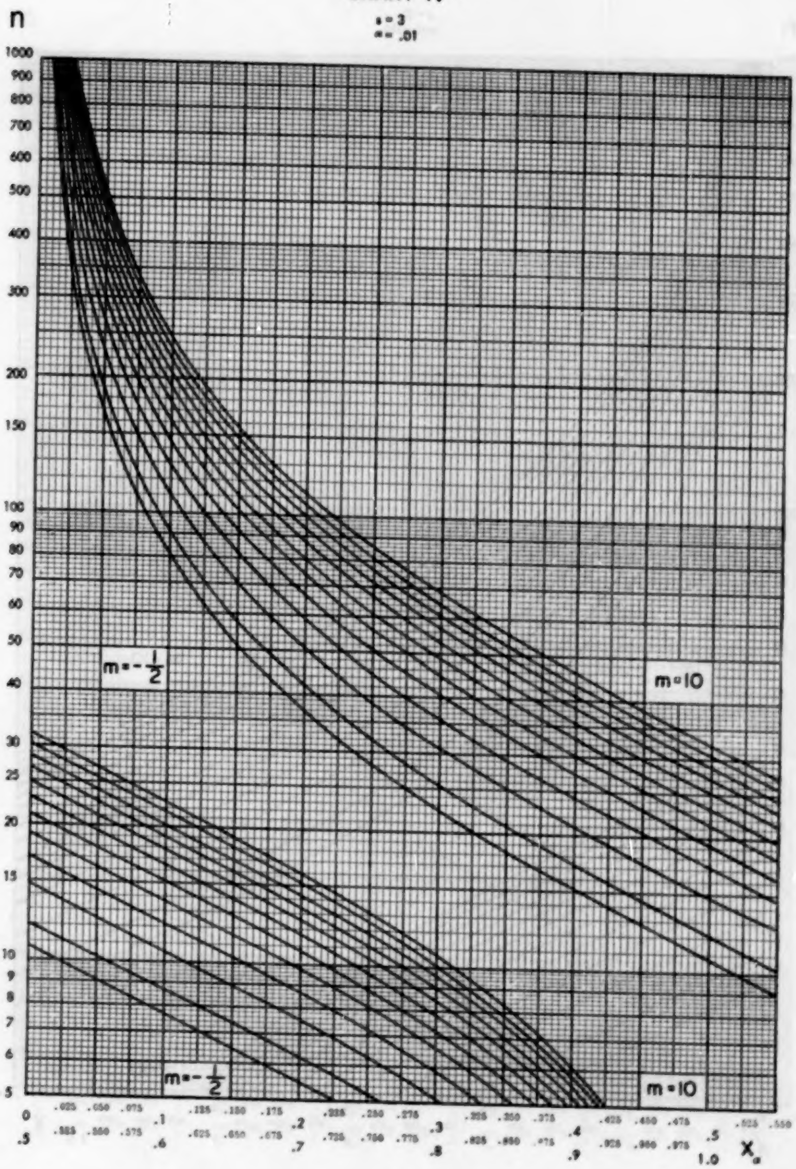


CHART V

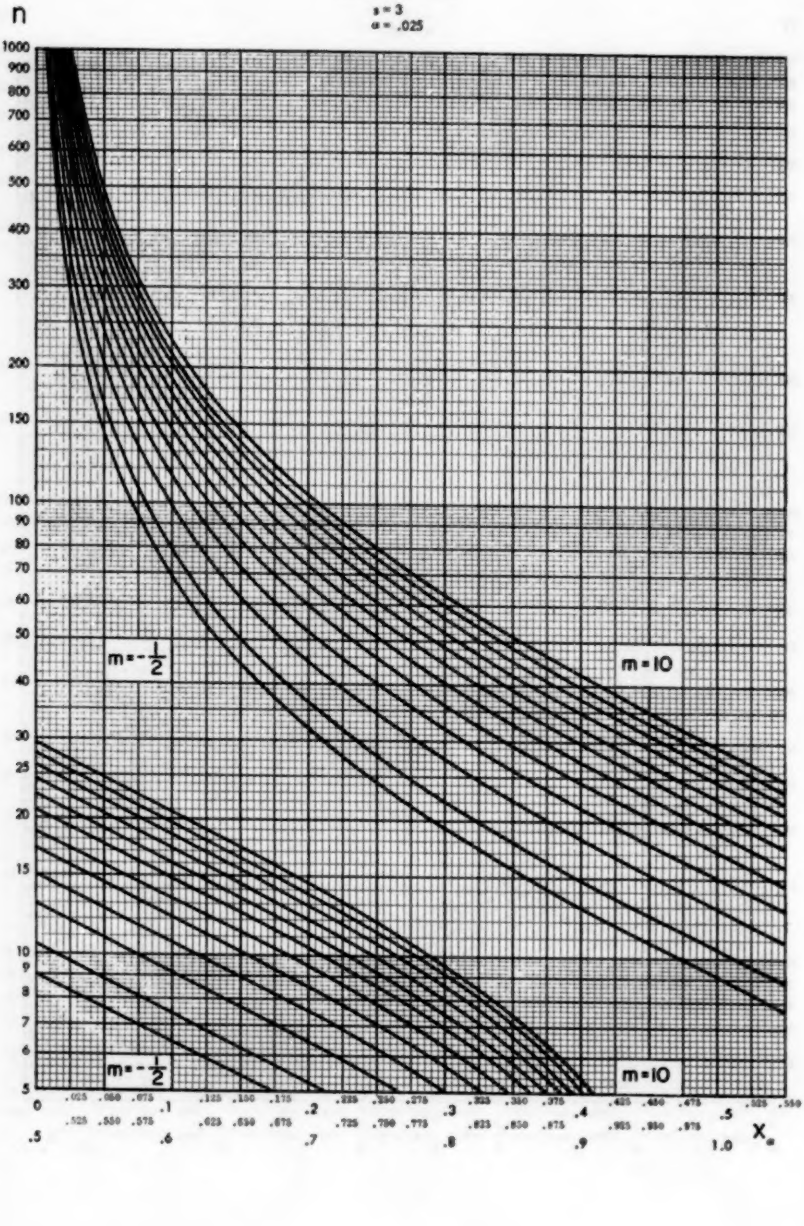

s = 3
α = .025
n
1000
900
800
700
600
500
400
300
200
150
100
90
80
70
60
50
40
30
25
20
15
10
9
8
7
6
5
m = -1/2
m = 10
m = -1/2
m = 10
0 .025 .050 .075 .1 .125 .150 .175 .2 .225 .250 .275 .3 .325 .350 .375 .4 .425 .450 .475 .5 .525 .550
.525 .550 .575 .625 .650 .675 .725 .750 .775 .825 .850 .875 .925 .950 .975
.5 .6 .7 .8 .9 1.0
X_α

CHART VI

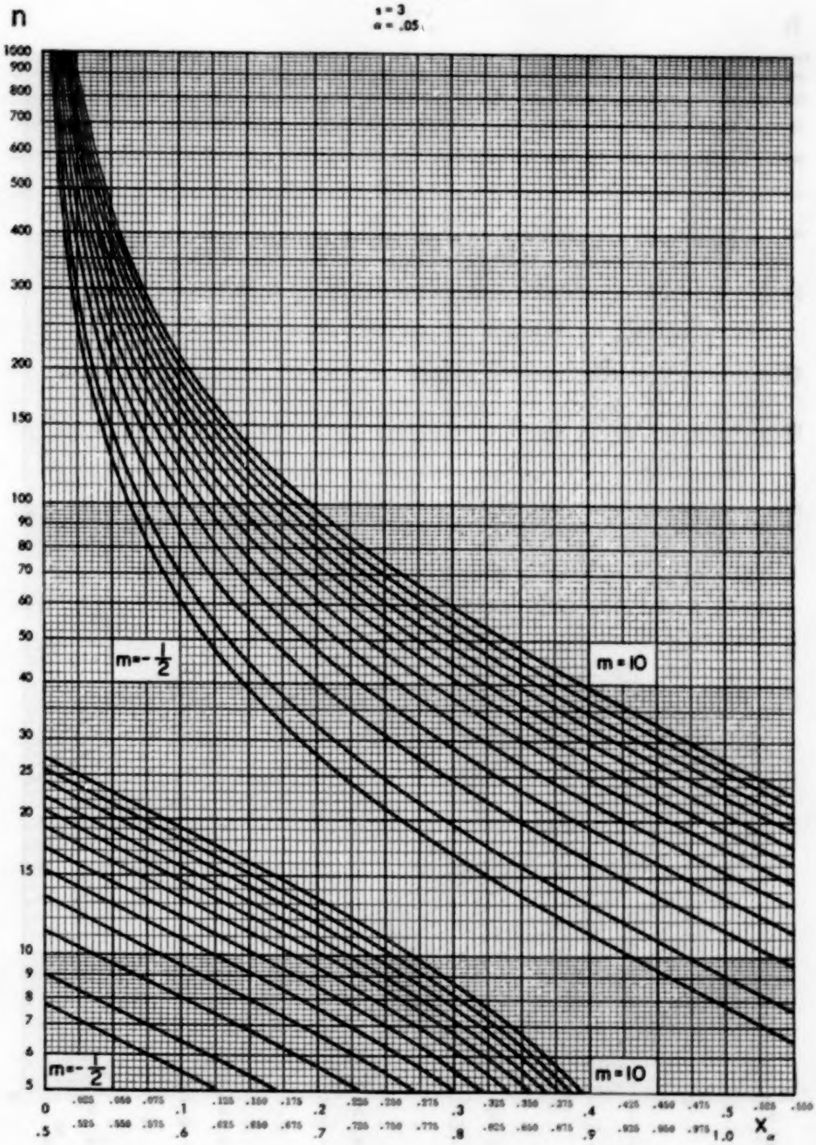


CHART VII

$s = 4$
$\alpha = .01$

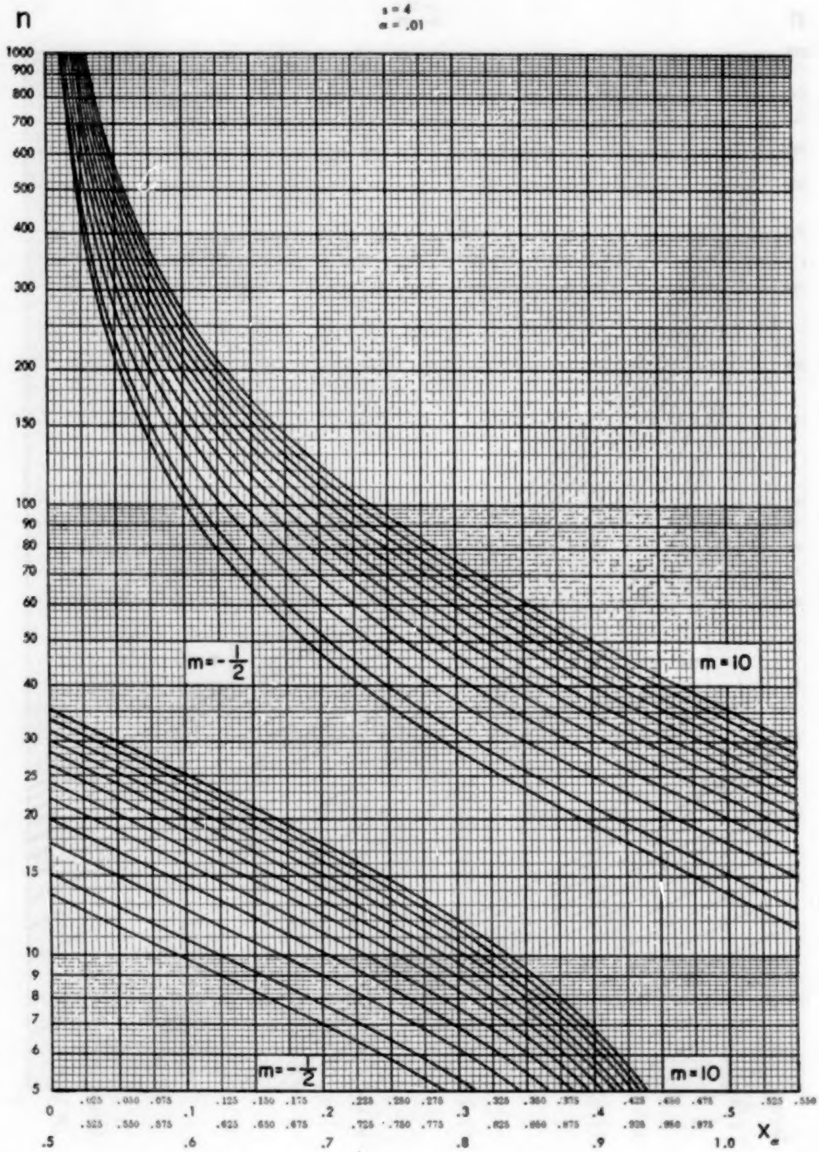


CHART VIII

s = 4
α = .025

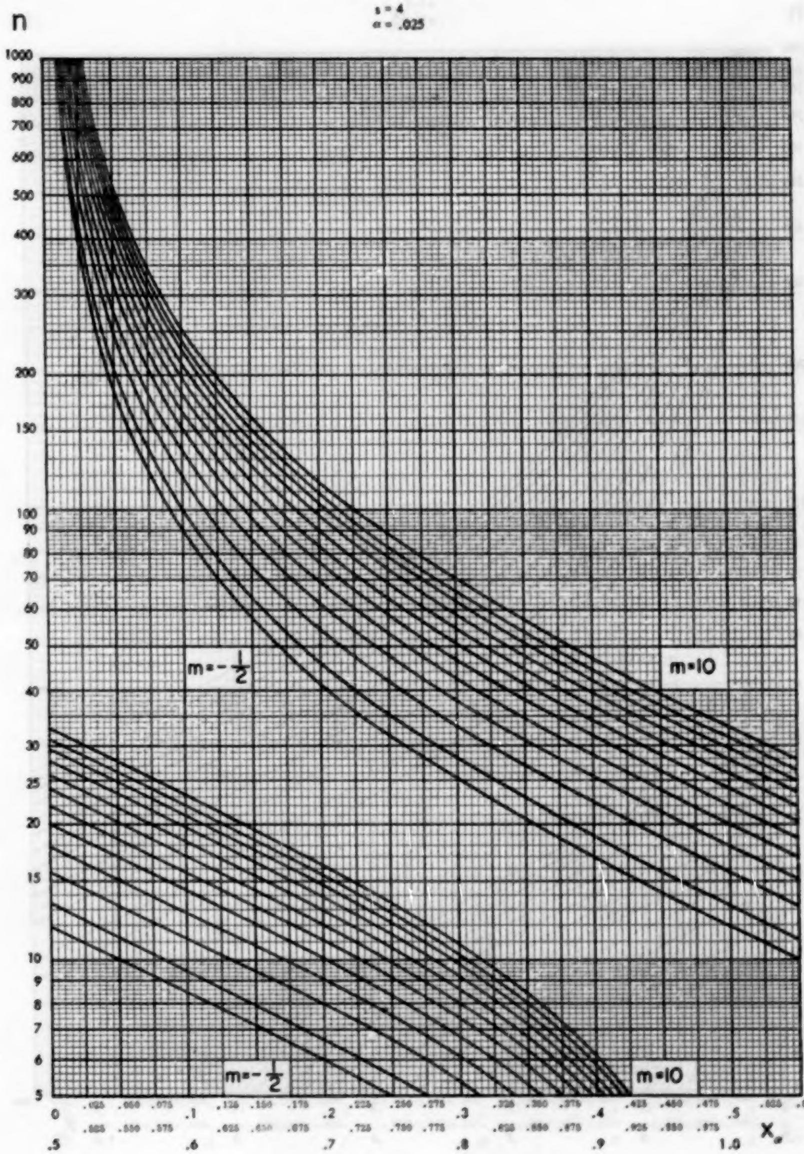


CHART IX

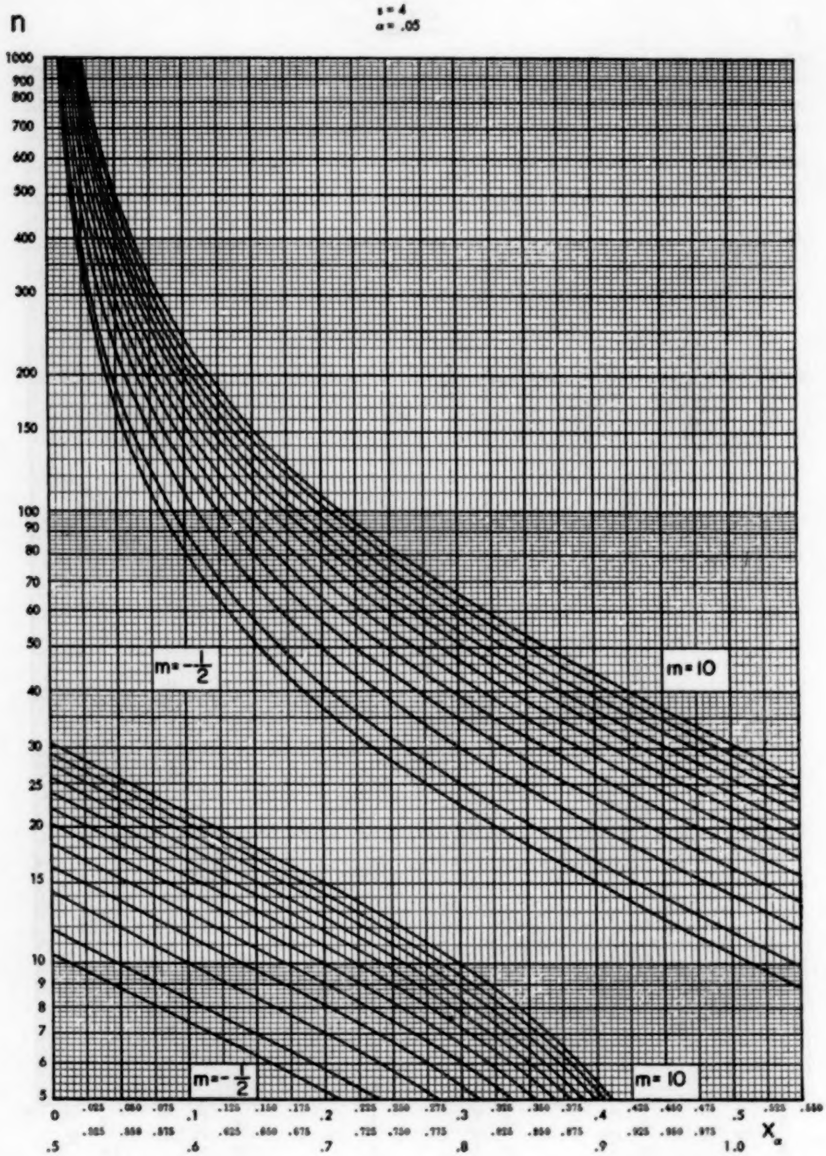


CHART X

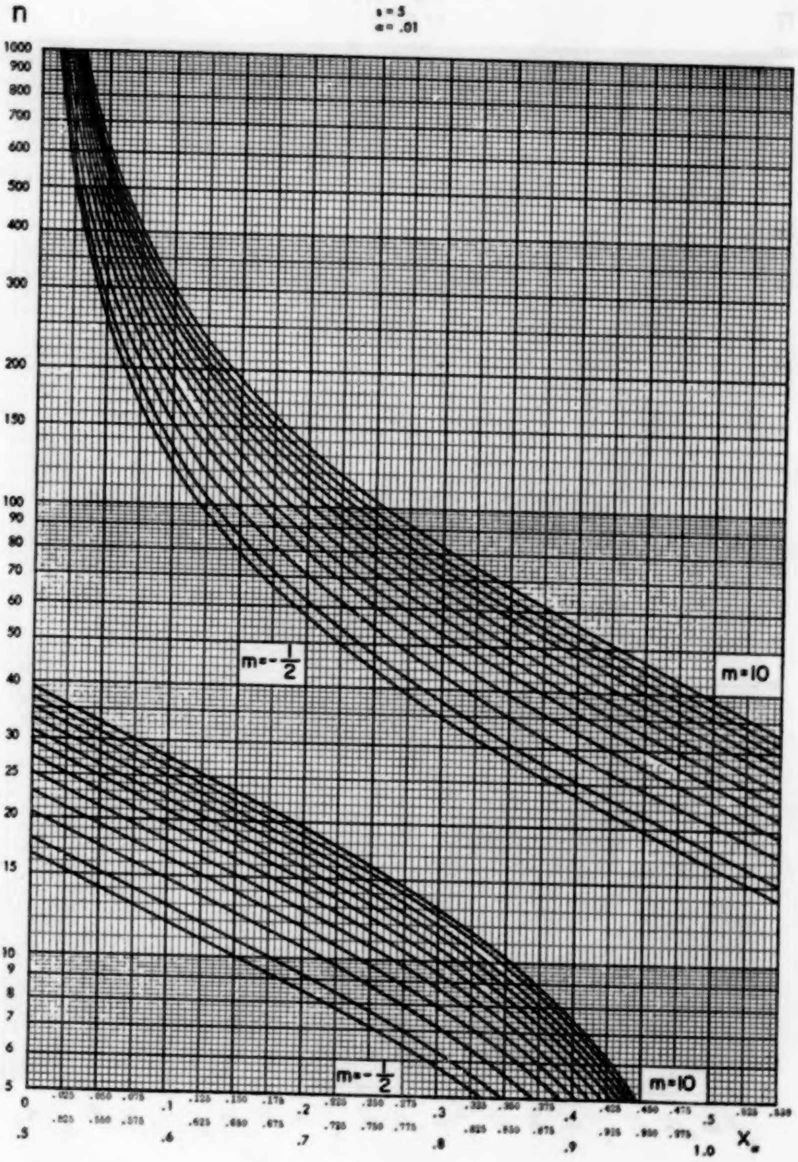


CHART XI

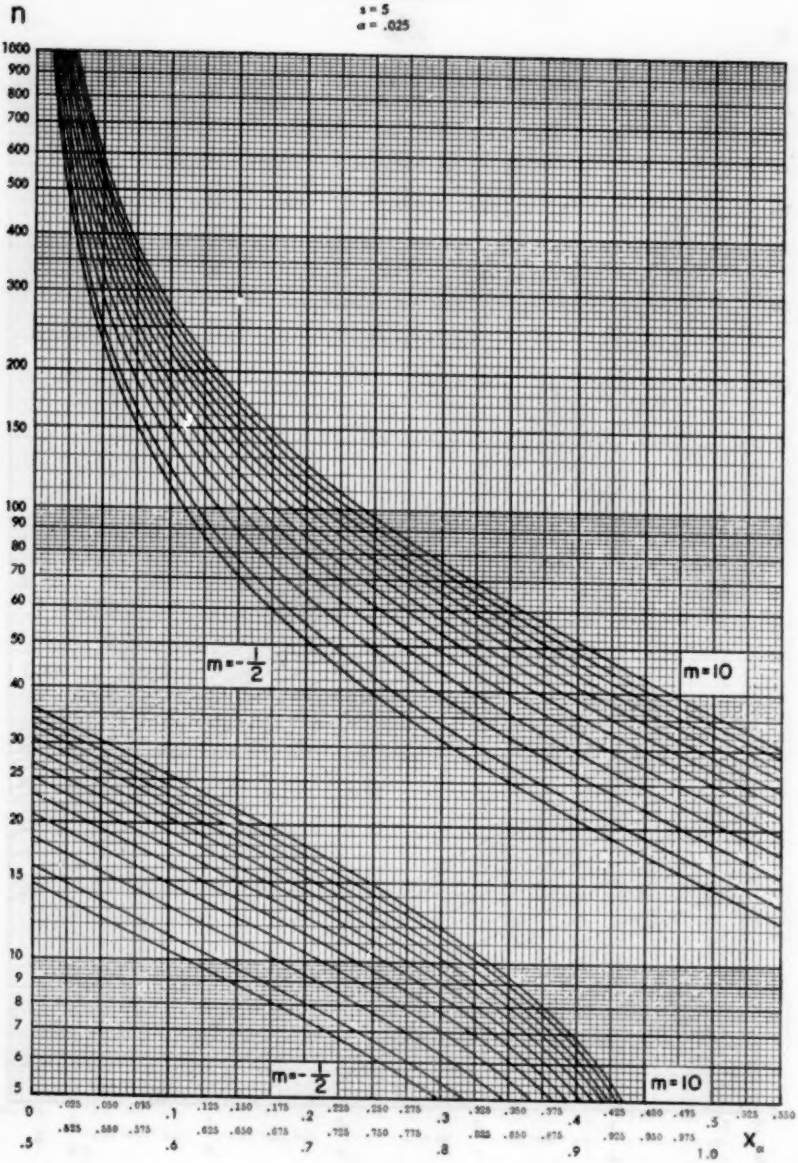
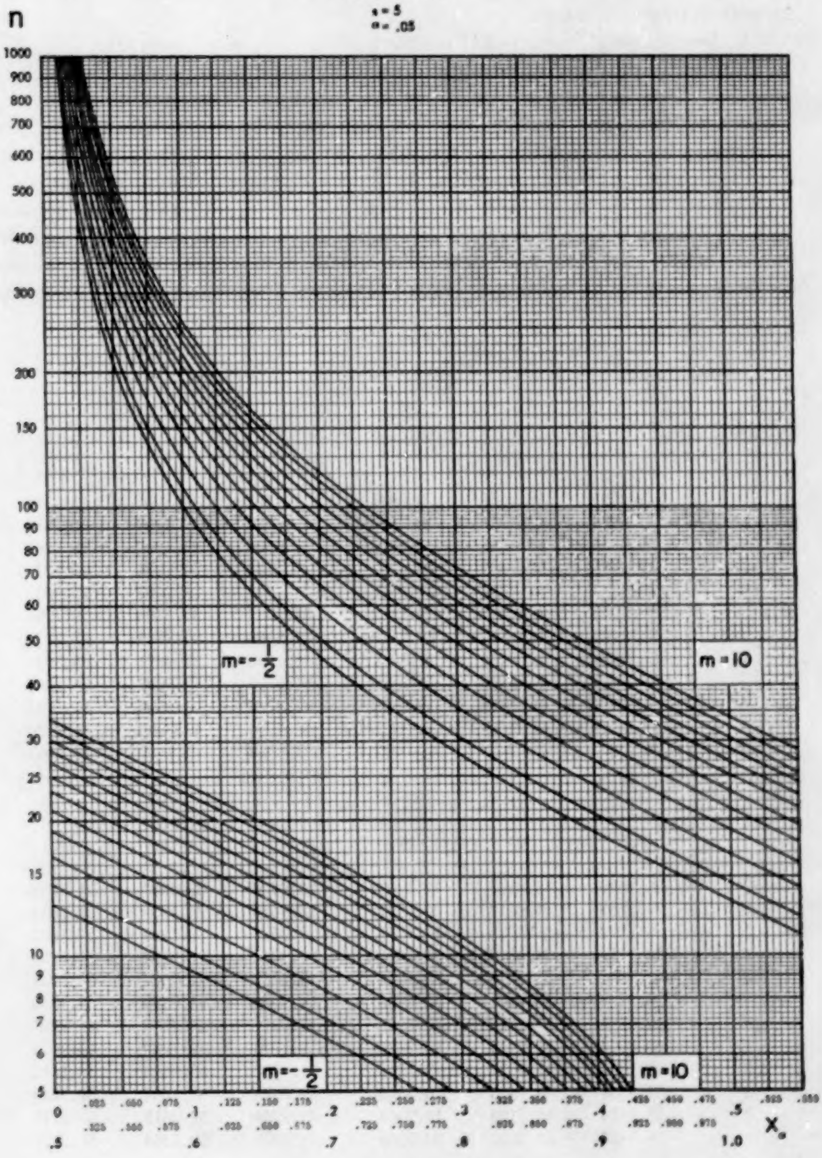


CHART XII

$s = 5$
$\alpha = .05$



3. Charts of the upper 1%, 2.5%, and 5% points of the distribution of the largest characteristic root.

3.1. *Description.* Charts I–XII enable finding $x_\alpha(s, m, n)$ such that

$$P[\theta_s \leqq x_\alpha(s, m, n)] = 1 - \alpha,$$

where θ_s is the largest non-zero root. On each page, the graphs appear for a particular s and α ($s = 2(1)5$, $\alpha = .01, .025, .05$) for $m = -\frac{1}{2}, 0(1)10$ and n from 5 to 1000. The curves corresponding to the twelve values of m on each page are in two sections, the lower section being the continuation of the upper section, with an overlap occurring from $x_\alpha = .50$ to $.55$. Of the two scales for x_α at the bottom of the page, the upper scale corresponds to the upper set of curves and the lower scale to the lower set. The lowest curve in each case (with the excep-

TABLE 4.1

Values of $z_\alpha(s, m)$

m \ α	$s = 2$			$s = 3$		
	.01	.025	.05	.01	.025	.05
$-\frac{1}{2}$	12.1601	10.1465	8.5941	17.1762	14.9006	13.1141
0	14.5680	12.4157	10.7393	19.5012	17.1192	15.2389
1	18.7346	16.3599	14.4873	23.6906	21.1262	19.0866
2	22.4664	19.9086	17.8762	27.5181	24.7971	22.6216
3	25.9526	23.2352	21.0641	31.1203	28.2597	25.9635
4	29.2755	26.4145	24.1192	34.5647	31.5768	29.1708
5	32.4795	29.4870	27.0779	37.8905	34.7848	32.2774
6	35.5920	32.4773	29.9628	41.1230	37.9071	35.3050
7	38.6311	35.4018	32.7886	44.2795	40.9597	38.2685
8	41.6098	38.2722	35.5658	47.3726	43.9542	41.1785
9	44.5375	41.0970	38.3021	50.4118	46.8993	44.0430
10	47.4215	43.8827	41.0033	53.4042	49.8017	46.8684

m \ α	$s = 4$			$s = 5$		
	.01	.025	.05	.01	.025	.05
$-\frac{1}{2}$	21.9646	19.4847	17.5183	26.6206	23.9697	21.8538
0	24.2395	21.6713	19.6277	28.8613	26.1339	23.9515
1	28.4328	25.7078	23.5278	33.0524	30.1861	27.8835
2	32.3175	29.4540	27.1543	36.9748	33.9834	31.5731
3	35.9964	33.0074	30.5996	40.7087	37.6027	35.0938
4	39.5253	36.4207	33.9135	44.3009	41.0883	38.4880
5	42.9387	39.7262	37.1265	47.7814	44.4688	41.7829
6	46.2593	42.9454	40.2588	51.1710	47.7639	44.9971
7	49.5034	46.0934	43.3246	54.4847	50.9876	48.1441
8	52.6831	49.1815	46.3345	57.7338	54.1508	51.2340
9	55.8073	52.2182	49.2964	60.9269	57.2615	54.2745
10	58.8833	55.2102	52.2166	64.0709	60.3264	57.2717

tion of Chart III) corresponds to $m = -\frac{1}{2}$, the next lowest to $m = 0$, the next to $m = 1$, etc., to the uppermost curve, which corresponds to $m = 10$. The scale for n is on the left margin of the page and is logarithmic.

3.2. *Note.* The values of $x_\alpha(s, m, n)$ may be read from the charts correct to two decimals. For a more precise value, when $n > 100$, the method described in Section 4 is suggested.

4. Asymptotic $z_\alpha(s, m)$ values.

4.1. *Description.* In Table 4.1 the values of $z_\alpha(s, m)$ are listed for $s = 2(1)5$, $m = -\frac{1}{2}, 0(1)10$, and $\alpha = .01, .025, .05$. For $n > 100$, these may be used to obtain $x_\alpha(s, m, n)$, with an error of at most five units in the fourth decimal. For a given combination (s, m, n) and a desired significance level α, determine $x = x_\alpha(s, m, n)$ from (2.1) with $z = z_\alpha(s, m)$ obtained from Table 4.1.

5. Acknowledgments. I should like to express my sincere thanks to S. N. Roy and R. E. Bargmann for their helpful advice and assistance in the preparation of this paper.

REFERENCES

[1] Hebron E. Adams, "DOPSIR: Double precision floating point SOAP interpretive routine," 650 Program Library, File No. 2.0.010, IBM, Washington, D.C. (1956).

[2] T. W. Anderson, *Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York, 1958.

[3] Rolf Bargmann, "A study of independence and dependence in multivariate normal analysis," North Carolina Institute of Statistics Mimeograph Series No. 186 (1957).

[4] M. S. Bartlett, "Further aspects of the theory of multiple regression," *Proc. Cambridge Phil. Soc.*, Vol. 34 (1938), pp. 33–40.

[5] S. B. Chaudhuri, "Statistical tables and certain recurrence relations connected with p-statistics," *Calcutta Stat. Assoc. Bull.*, Vol. 6 (1956), pp. 181–188.

[6] R. A. Fisher, "The sampling distribution of some statistics obtained from non-linear equations," *Ann. Eugenics*, Vol. 9 (1939), pp. 238–249.

[7] F. G. Foster and D. H. Rees, "Upper percentage points of the generalized beta distribution. I," *Biometrika*, Vol. 44 (1957), pp. 237–247.

[8] F. G. Foster, "Upper percentage points of the generalized beta distribution. II," *Biometrika*, Vol. 44 (1957), pp. 441–453.

[9] F. G. Foster, "Upper percentage points of the generalized beta distribution. III," *Biometrika*, Vol. 45 (1958), pp. 492–503.

[10] M. A. Girshick, "On the sampling theory of the roots of determinantal equations," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 203–224.

[11] P. L. Hsu, "On the distribution of roots of certain determinantal equations," *Ann. Eugenics*, Vol. 9 (1939), pp. 250–258.

[12] D. N. Nanda, "Distribution of a root of a determinantal equation," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 47–57.

[13] D. N. Nanda, "Probability distribution tables of the larger root of a determinantal equation with two roots," *J. Indian Soc. of Agricultural Stat.*, Vol. 3 (1951), pp. 175–177.

[14] K. C. Sreedharan Pillai, "On the distribution of the largest or the smallest root of a matrix in multivariate analysis," *Biometrika*, Vol. 43 (1956), pp. 122–127.

[15] K. C. Sreedharan Pillai, *Concise Tables for Statisticians*, The Statistical Center University of the Philippines, Manila, 1957.

[16] K. C. Sreedharan Pillai and Celia G. Bantegui, "On the distribution of the largest of six roots of a matrix in multivariate analysis," *Biometrika*, Vol. 46 (1959), pp. 237–240.

[17] C. Radhakrishna Rao, *Advanced Statistical Methods in Biometric Research*, John Wiley and Sons, New York, 1952.

[18] S. N. Roy, "p-statistics or some generalizations in analysis of variance appropriate to multivariate problems," *Sankhyā*, Vol. 4 (1939), pp. 381–396.

[19] Samarendra Nath Roy, "The individual sampling distribution of the maximum, the minimum and any intermediate of the p-statistics on the null hypothesis," *Sankhyā*, Vol. 7 (1945), pp. 133–158.

[20] S. N. Roy, "On a heuristic method of test construction and its use in multivariate analysis," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 220–238.

[21] S. N. Roy and R. C. Bose, "Simultaneous confidence interval estimation," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 513–536.

[22] S. N. Roy, "Some further results in simultaneous confidence interval estimation," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 752–761.

[23] S. N. Roy, *Some Aspects of Multivariate Analysis*, John Wiley and Sons, New York, 1957.

[24] S. N. Roy and R. Gnanadesikan, "Further contributions to multivariate confidence bounds," *Biometrika*, Vol. 44 (1957), pp. 399–410.

[25] S. N. Roy and R. E. Bargmann, "Tests of multiple independence and the associated confidence bounds," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 491–503.

[26] S. N. Roy and R. F. Potthoff, "Confidence bounds on vector analogues of the 'ratio of means' and the 'ratio of variances' for two correlated normal variates and some associated tests," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 829–841.

[27] S. N. Roy and J. Roy, "Analysis of variance with univariate or multivariate, fixed or mixed classical models," North Carolina Institute of Statistics Mimeograph Series No. 208 (1958).

[28] John R. B. Whittlesey, "Summary of results obtained under the Ford Foundation contract," unpublished research report, Dept. of Statistics, University of North Carolina (1954).

[29] S. S. Wilks, "Certain generalizations in the analysis of variance," *Biometrika*, Vol. 24 (1932), pp. 471–494.

[30] S. S. Wilks, *Mathematical Statistics*, Princeton, N. J., Princeton Univ. Press, 1943.

[31] V. M. Wolontis, "A complete floating-decimal interpretive system for the IBM 650 magnetic drum calculator," IBM Applied Science Div., *Tech. Newsletter No. 11* (1956).

CORRECTION FOR BIAS INTRODUCED BY A TRANSFORMATION OF VARIABLES[1]

BY JERZY NEYMAN AND ELIZABETH L. SCOTT

University of California, Berkeley

1. Introduction. The problem of "normalizing" transformations has two different facets: one is concerned with the identity of transforming functions suitable for variables following a distribution with a particular shape or properties and the other with the nature of the statistics capable of serving as unbiased estimates in cases where a given transforming function appears to be successful. The literature on the first problem is rich (see, for example, [1], [2] and [8]). This paper is concerned with the second problem. Our purpose is to deduce minimum variance unbiased estimates of the effects of experimental treatments expressed in the original units. The solution is obtained for a broad category of transforming functions.

The estimates of treatment effects expressed in the original units are customarily obtained by the inverse transformation of the estimates in transformed units. As is well known [1], [6], [7], this traditional estimate is biased. Occasionally, this bias is important. Further, the bias gains importance when a number of similar estimates of the same effect, obtained from independent sets of observations, are averaged in order to estimate the average effect. The random errors of the particular effects tend to average out but, in general, not the bias.

2. Statement of the problem. Our basic assumption in this paper is that the transformation used in the analysis of an experiment is faultless so that the transformed variables exactly follow normal distributions with some postulated means and with the same unknown variance σ^2. Generically, these normal variables will be denoted by the letter $\xi(\psi)$ where ψ identifies the expectation of the variable concerned. Thus $\xi(\psi)$ is the transformed variable in the experiment. The variable that is directly observable will be denoted by $X(\psi)$. It will be assumed that

$$X(\psi) = f[\xi(\psi)], \tag{1}$$

where f is a strictly increasing function defined for all real values of its argument. Later on, we shall introduce further limitations on f. It will be noticed that f is the inverse of the function used for transforming the observable variable X. For example, with the square root transformation the function f is the square of its argument.

The problem treated is concerned with a particular pair of variables of the

Received August 8, 1959; revised February 26, 1960.

[1] This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-43). This paper in whole or in part may be reproduced for any purpose of the United States Government.

family considered, namely with $\xi(\mu)$ and $X(\mu)$, where μ is the mean of $\xi(\mu)$ and is a well-defined but unknown number. Specifically, we are concerned with estimating

$$\theta = E[X(\mu)]. \tag{2}$$

Our problem arises when the variables $\xi(\mu)$ and $X(\mu)$ are not directly observable. On the other hand, the variables that are observable in the given experiment yield a pair of statistics, $\hat{\mu}$ and S^2, mutually independent and jointly sufficient for μ and σ^2. The first is a normal variable with mean μ and variance $\lambda^2\sigma^2$, where λ^2 is a known number. The second statistic, S^2, is the residual sum of squares and, divided by σ^2, is distributed as χ^2 with a certain number ν of degrees of freedom. Our problem is to devise a function, say $\hat{\theta}(\hat{\mu}, S^2)$ such that its expectation equals θ identically in μ and σ^2. Because of the familiar result of Lehmann and Scheffé [5] that the sufficient system of statistics $(\hat{\mu}, S^2)$ is boundedly complete, it follows that the function $\hat{\theta}(\hat{\mu}, S^2)$ is unique and is the minimum variance unbiased estimate of θ.

Before proceeding to the construction of the estimate $\hat{\theta}(\hat{\mu}, S^2)$ we give two illustrative examples.

3. Example 1: An experiment in randomized blocks. Denote by α and β two unknown parameters capable of assuming values within a certain open set, and by $\xi(\alpha, \beta)$ a normal random variable with expectation

$$E[\xi(\alpha, \beta)] = \alpha + \beta \tag{3}$$

and with a fixed variance σ^2. Correspondingly,

$$X(\alpha, \beta) = f[\xi(\alpha, \beta)]. \tag{4}$$

A randomized block experiment will yield particular values of mn independent random variables $X(\alpha_i, \beta_j)$ for $i = 1, 2, \cdots, m$ and $j = 1, 2, \cdots, n$, with $\sum \beta_j = 0$. Here the β's represent the familiar block effects and α_i stands for the "transformed effect" of the ith treatment in the hypothetical average conditions of the experiment. The analysis of the experiment ordinarily involves the estimation in original units (pounds, inches, number of surviving insects, etc.) of the effect of the ith treatment if it were applied in the average conditions of the experiment. In order that this estimate can be conveniently combined, by averaging, with similar estimates derived from other experiments involving the same treatment, the estimate sought should be unbiased. The quantity to estimate[2] is, then,

$$\theta = E[X(\alpha_i, 0)] = E\{f[\xi(\alpha_i, 0)]\} \tag{5}$$

where $\xi(\alpha_i, 0)$ is a normal variable with an unknown mean $\alpha_i = \mu$ and with

[2] Of course, the definition of the "effect of the ith treatment in the average conditions of the experiment" by means of Formula (5) is not the only possible definition of this concept. An alternative definition might be the average over j of the quantities $E\{f[\xi(\alpha_i, \beta_j)]\}$.

variance σ^2. While $\xi(\alpha_i, 0)$ and $X(\alpha_i, 0)$ are not directly observable, the experiment yields an estimate of μ, namely,

$$\hat{\mu} = \frac{1}{n} \sum_{j=1}^{n} \xi_{ij}, \tag{6}$$

which, according to traditional theory, is normal with mean $\mu = \alpha_i$ and variance

$$\sigma_{\hat{\mu}}^2 = \lambda^2\sigma^2 = \sigma^2/n. \tag{7}$$

Also, in the usual notation,

$$S^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} (\xi_{ij} - \xi_{i.} - \xi_{.j} + \xi_{..})^2 \tag{8}$$

is the sum of squares of the residuals which, combined with $\hat{\mu}$, forms a sufficient system of statistics for μ and σ^2. It is independent of $\hat{\mu}$, and is distributed as the product of σ^2 by a χ^2 with $\nu = (m - 1)(n - 1)$ degrees of freedom. Our problem is to devise a function, $\hat{\theta}(\hat{\mu}, S^2)$, which is an unbiased estimate of θ.

4. Example 2: Regression analysis of a randomized cloud seeding experiment. An experiment is performed to check whether the "seeding" of clouds, intended to increase the precipitation in a "target area" T, has an effect. Also, it is intended to estimate the amount of this effect measured in inches of actual precipitation. A certain number $s \geqq 1$ of adjoining areas, presumed to be unaffected by seeding, are used as controls. We shall use the symbol X_i to denote the rainfall from a particular storm falling in the ith control area and the vector symbol $X = (X_1, X_2, \cdots, X_s)$ to denote the precipitation from the same storm in all the controls. For each X we consider the random variable $Y(X)$ representing the target precipitation in conditions when the precipitation in the controls is X and there is no seeding. All these variables are measured in inches.

Now suppose that a storm, with control precipitation equal to X', is seeded and yields Y' inches of rain in the target. In order to estimate the effect of this seeding it is necessary to have an estimate of the rain which would have fallen in the target from the same storm if there were no seeding. In other words, we need an estimate of $E[Y(X')] = \theta$. The hypotheses usually made about the variables X and $Y(X)$ are that, by means of some suitable change of scale, $X_i = f(\xi_i)$, etc., they can be replaced by transformed variables ξ and $\eta(\xi)$, respectively, such that, for each ξ, the variable $\eta(\xi)$ is (approximately) normally distributed with a mean

$$E[\eta(\xi)] = \alpha_0 + \sum_{i=1}^{s} \alpha_i \xi_i, \tag{9}$$

where the α's are unknown constants, and with a variance σ^2 independent of ξ. With these assumptions, the quantity θ to be estimated is

$$\theta = E\{f[\eta(\xi')]\}, \tag{10}$$

where ξ' stands for the transformed value of X'. Denote by μ the expectation of $\eta(\xi')$,

$$\mu = \alpha_0 + \sum_{i=1}^{s} \alpha_i \xi_i' , \tag{11}$$

and by $\hat{\mu}$ its minimum variance unbiased estimate obtained from the regression analysis of a random sample of unseeded storms. The same analysis provides the sum S^2 of squares of residuals, which is stochastically independent of $\hat{\mu}$ and, when divided by σ^2, is distributed as χ^2 with a certain number of degrees of freedom ν. The variance of $\hat{\mu}$ is $\lambda^2(\xi')\sigma^2$, where and factor $\lambda^2(\xi')$ depends upon the value of ξ' and, in fact, grows without limit when ξ' diverges from the average $\bar{\xi}$ of the control precipitation from the nonseeded storms used to evaluate $\hat{\mu}$.

Our problem consists in determining a function $\hat{\theta}(\hat{\mu}, S^2)$ such that

$$E[\hat{\theta}(\hat{\mu}, S^2)] = \theta. \tag{12}$$

The difference between Y' and $\hat{\theta}(\hat{\mu}, S^2)$ is the estimated effect of seeding, expressed in inches.

5. Method and auxiliary formulas. Since the normalizing transformations are supposed to amount to a change in scale of measuring the observable random variables, it is natural to assume that the function f determining the observable random variable X in terms of the normal variable ξ is fairly regular. Our method, to be termed the expansion method, of constructing $\hat{\theta}(\hat{\mu}, S^2)$ is limited to the case where (i) $\theta = E\{f[\xi(\mu)]\}$ exists, (ii) f is an entire function, and (iii) the expectation θ may be obtained by taking expectations, term by term, of the Taylor expansion of f, so that

$$\theta = E\{f[\xi(\mu)]\} = f(0) + \sum_{n=1}^{\infty} \frac{1}{n!} f^{(n)} E[\xi^n(\mu)], \tag{13}$$

where $f^{(n)}$ stands for the nth derivative of f evaluated at zero. Then, for each n, we determine a homogeneous combination

$$T_n = \sum_{k=0}^{n} A_{n,k} \hat{\mu}^k S^{n-k} \tag{14}$$

such that

$$E(T_n) = E[\xi^n(\mu)], \tag{15}$$

and show that

$$\hat{\theta}(\hat{\mu}, S^2) = f(0) + \sum_{n=1}^{\infty} \frac{1}{n!} f^{(n)} T_n \tag{16}$$

is the solution of the problem.

Also, for functions f of a particular family, we give an alternative easy method of constructing $\hat{\theta}(\hat{\mu}, S^2)$. Before proceeding we must recall certain formulas and deduce certain bounds.

For every $m = 1, 2, \cdots$, we have

$$E[S^{2m}] = (2\sigma^2)^m \Gamma(\tfrac{1}{2}\nu + m)/\Gamma(\tfrac{1}{2}\nu). \tag{17}$$

Also

$$E[\xi^{2m}(\mu)] = \sum_{k=0}^{m} \frac{(2m)!}{(2k)!(m-k)!} \mu^{2k}(\sigma^2/2)^{m-k}, \tag{18}$$

and it follows that

$$\frac{(2m)!}{m!}(\sigma^2/2)^m \leqq E[\xi^{2m}(\mu)] \leqq \frac{(2m)!}{m!}[(\mu^2+\sigma^2)/2]^m. \tag{19}$$

Similarly

$$E[\xi^{2m+1}(\mu)] = \mu \sum_{k=0}^{m} \frac{(2m+1)!}{(2k+1)!(m-k)!} \mu^{2k}(\sigma^2/2)^{m-k}. \tag{20}$$

In order to obtain convenient bounds on $E(|\xi^{2m+1}|)$, we notice first that

$$|\mu| \frac{(2m+1)!}{m!}(\sigma^2/2)^m \leqq |E[\xi^{2m+1}(\mu)]| < E|\xi^{2m+1}(\mu)|. \tag{21}$$

Further, by Schwarz' inequality and because of (19),

$$\begin{aligned} E|\xi^{2m+1}(\mu)| &< \{E[\xi^2(\mu)]E[\xi^{4m}(\mu)]\}^{\frac{1}{2}} \\ &\leqq (\mu^2+\sigma^2)^{\frac{1}{2}}\left(\frac{(4m)!}{(2m)!}\right)^{\frac{1}{2}}[(\mu^2+\sigma^2)/2]^m. \end{aligned} \tag{22}$$

However, it is easy to see that

$$\frac{m!}{2^m(2m+1)!}\left(\frac{(4m)!}{(2m)!}\right)^{\frac{1}{2}} = \left\{\prod_{k=1}^{m} \frac{(4k-3)(4k-1)}{(4k+2)^2}\right\}^{\frac{1}{2}} < 1. \tag{23}$$

Consequently, we may write

$$|\mu| \frac{(2m+1)!}{m!}(\sigma^2/2)^m < E|\xi^{2m+1}(\mu)| < \frac{(2m+1)!}{m!}[\mu^2+\sigma^2]^m, \tag{24}$$

for all m.

6. Term by term evaluation of the expectation of a Taylor series. In this section we use the bounds found in Section 5 in order to prove certain theorems.

THEOREM 1. *In order that the series in the right-hand side of* (13) *be convergent irrespective of the values of* μ *and* σ^2, *it is necessary and sufficient that the radii of convergence of the two series*

$$\sum_{n=0}^{\infty} \frac{1}{n!} f^{(2n)} z^n \quad \text{and} \quad \sum_{n=0}^{\infty} \frac{1}{n!} f^{(2n+1)} z^n \tag{25}$$

both be infinite, so that

$$
\begin{aligned}
(26)\qquad \lim_{n\to\infty}\left(\frac{1}{n!}\,|f^{(2n)}|\right)^{1/n} &= \lim_{n\to\infty}\left(\frac{1}{n!}\,|f^{(2n+1)}|\right)^{1/n} \\
&= \lim_{n\to\infty}\frac{1}{n}\,(|f^{(2n)}|)^{1/n} = \lim_{n\to\infty}\frac{1}{n}\,(|f^{(2n+1)}|)^{1/n} = 0.
\end{aligned}
$$

It will be noticed that conditions (26) are stronger than the assumption that the Taylor expansion of f,

$$
(27)\qquad f(\xi) = \sum_{n=0}^{\infty}\frac{1}{n!}f^{(n)}\xi^n,
$$

is convergent for all real ξ. In fact, the conditions necessary and sufficient for this to happen may be written as

$$
(28)\qquad \lim_{n\to\infty}\frac{1}{n^2}\,(f^{(2n)})^{1/n} = \lim_{n\to\infty}\frac{1}{n^2}\,(f^{(2n+1)})^{1/n} = 0.
$$

Thus, if the radii of convergence of (25) are infinite, then that of (27) will also be infinite, but the converse is not necessarily true.

In order to prove the therem we simply notice that inequalities (19) and (24) imply that, for all m,

$$
(29)\qquad \frac{1}{m!}\,|f^{(2m)}|\,(\sigma^2/2)^m \leqq \frac{1}{(2m)!}\,|f^{(2m)}|\,E[\xi^{2m}(\mu)] \leqq \frac{1}{m!}\,|f^{(2m)}|\,[(\mu^2+\sigma^2)/2]^m,
$$

$$
\begin{aligned}
(30)\qquad \frac{1}{m!}\,|\mu f^{(2m+1)}|\,(\sigma^2/2)^m &\leqq \frac{1}{(2m+1)!}\,|f^{(2m+1)}E[\xi^{2m+1}(\mu)]| \\
&\leqq \frac{1}{m!}\,|f^{(2m+1)}|\,[\mu^2+\sigma^2]^m.
\end{aligned}
$$

If we assume that the series (13) is convergent for all values of μ and σ^2, then this will imply that the middle terms in (29) and (30) tend to zero as $m \to \infty$. In turn, this implies that, for all σ^2 and for sufficiently large m,

$$
(31)\qquad \left(\frac{1}{m!}\,|f^{(2m)}|\right)^{1/m} < \frac{2}{\sigma^2}, \qquad \left(\frac{1}{m!}\,|f^{(2m+1)}|\right)^{1/m} < \frac{2}{\sigma^2},
$$

which is equivalent to (26). On the other hand, if we assume that conditions (26) are satisfied, then the two series (25) are absolutely convergent for all values of the argument and the right inequalities (29) and (30) imply absolute convergence of (13).

THEOREM 2. *Under the conditions of Theorem* 1, *that is, under conditions* (26),

$$
(32)\qquad \theta = E\{f[\xi(\mu)]\} = f(0) + \sum_{n=1}^{\infty}\frac{1}{n!}f^{(n)}E[\xi^n(\mu)]
$$

for all μ and σ^2.

In other words, if conditions (26) are satisfied then the expectation of f can be obtained by taking expectations term by term of the Taylor expansion of this function.

Theorem 2 is implied by inequality (24) showing that in the middle part of Formula (30) the expectation of $\xi^{2m+1}(\mu)$ may be replaced by the expectation of the absolute value $|\xi^{2m+1}(\mu)|$.

For convenience of reference we shall adopt the following definition.

DEFINITION. *An entire function f is called of second order if it satisfies conditions* (26).

It will be seen that every indefinitely differentiable function whose derivatives at a particular point are bounded is necessarily a second order entire function. Also, the sum of two second order entire functions is itself a second order entire function.

7. Lack of complete generality of the expansion method. At this point it may be interesting to indicate a purely mathematical formulation of the general problem treated in this paper. This is as follows

For a given positive integer ν, for a given positive number λ^2 and for a given function f, defined on the real line and such that

$$\int_{-\infty}^{+\infty} |f(x)|\, e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}\, dx < +\infty \tag{33}$$

for all real μ and for $\sigma > 0$, to determine a function of two arguments $\theta(x, y^2)$, independent of μ and σ, such that

$$\begin{aligned}\lambda 2^{(\nu-2)/2}\sigma^\nu\Gamma(\tfrac{1}{2}\nu) \int_{-\infty}^{+\infty} f(x)e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}\, dx \\ = \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x-\mu)^2/\lambda^2\sigma^2} \int_0^\infty y^{\nu-1}e^{-\frac{1}{2}y^2/\sigma^2}\theta(x, y^2)\, dy\, dx,\end{aligned} \tag{34}$$

identically in μ and $\sigma > 0$.

The expansion method provides the solution of this problem when the function f is entire of second order. However, it is easy to construct entire functions f satisfying (33) which are not of the second order. One example is $\exp\{-x^2\}$. To such functions the expansion method is not applicable and we are not certain whether the solution of equation (34) exists.

8. Minimum variance unbiased estimate of the expectation of a second order entire function. From now on we shall deal exclusively with functions $f(\xi)$ which are entire of the second order.

Let $\hat{\mu}$ be a normal variable with expectation μ and variance $\lambda^2\sigma^2$ where λ^2 is a known number. Also, let S^2 be independent of $\hat{\mu}$ and such that S^2/σ^2 is distributed as χ^2 with ν degrees of freedom. Finally, for $n = 0, 1, 2, \cdots$,

$$T_{2n} = \sum_{k=0}^{n} \frac{(2n)!}{(2k)!(n-k)!}\, \hat{\mu}^{2k}[\tfrac{1}{4}S^2(1-\lambda^2)]^{n-k}\, \frac{\Gamma(\frac{1}{2}\nu)}{\Gamma(\frac{1}{2}\nu + n - k)} \tag{35}$$

and

$$(36)\quad T_{2n+1} = \sum_{k=0}^{n} \frac{(2n+1)!}{(2k+1)!(n-k)!}\, \hat{\mu}^{2k+1}[\tfrac{1}{4}S^2(1-\lambda^2)]^{n-k}\, \frac{\Gamma(\frac{1}{2}\nu)}{\Gamma(\frac{1}{2}\nu+n-k)}.$$

By direct computation it is easy to verify that, for every $m = 1, 2, \cdots,$

$$(37)\quad E(T_m) = E[\xi^m(\mu)].$$

Theorem 3. *If f is a second order entire function, then*

$$(38)\quad \theta(\hat{\mu}, S^2) = f(0) + \sum_{m=1}^{\infty} \frac{1}{n!} f^{(m)} T_m$$

is convergent for all values of $\hat{\mu}$ and S^2 and is an unbiased estimate of $\theta = E\{f[\xi(\mu)]\}$.

Comparing (35) and (36) with (18) and (20), noticing that

$$(39)\quad \Gamma(\tfrac{1}{2}\nu)/\Gamma(\tfrac{1}{2}\nu+n-k) \leqq (2/\nu)^{n-k}$$

and referring to (19), we find that

$$(40)\quad |T_{2n}| < \frac{(2n)!}{n!} Y^n$$

and, in a similar manner,

$$(41)\quad |T_{2n+1}| < |\hat{\mu}| \frac{2n+1)!}{n!} Y^n,$$

with

$$(42)\quad Y = [\nu\hat{\mu}^2 + S^2(1+\lambda^2)]/2\nu.$$

Because f is a second order entire function, it follows that the series (38) is absolutely convergent for all values of $\hat{\mu}$ and S^2. In order to prove that the expectation of $\hat{\theta}(\hat{\mu}, S^2)$ as defined by (38) can be obtained by taking expectations term by term, it is sufficient to show the convergence of the series obtained from (38) by replacing each T_m by the expectation of its absolute value. This is easily accomplished by noticing that $|T_m|$ cannot exceed the expression obtained by replacing in formulas (35) and (36) the value of $\hat{\mu}$ by that of $|\hat{\mu}|$ and $1 - \lambda^2$, which may be negative, by $1 + \lambda^2$. Further computations, similar to those leading to (19) and (24), indicate then that

$$(43)\quad \sum_{n=1}^{\infty} \frac{1}{n!} |f^{(n)}|\, E\,|T_n| < +\infty.$$

Because of (37), it follows that $\hat{\theta}(\hat{\mu}, S^2)$ as defined by (38) has the desired property of being an unbiased estimate of θ.

Formula (38) has the advantage of, so to speak, exhausting the method; it provides the minimum variance unbiased estimate of θ whatever the second order entire function f may be. This generality is paid for by the complexity of

the solution provided by (38). In the next section we give a somewhat simpler formula for $\hat{\theta}(\hat{\mu}, S^2)$ which is applicable when the function f satisfies a certain differential equation.

9. Alternative solution applicable to recursive type second order entire functions. We shall say that the function f is of recursive type if it satisfies the second order differential equation

$$f''(x) = A + Bf(x), \tag{44}$$

where A and B are arbitrary constants. However, in order to eliminate the trivial case where f is linear, we shall assume that at least one of the constants differs from zero. It is easy to verify that every recursive function is necessarily a second order entire function. This section will be limited to consideration of recursive type functions f. We shall be particularly interested in the expectation of their Taylor expansion about the point μ. Because the odd central moments of the normal variable are all equal to zero, we shall be concerned only with the derivatives of f of even order. We have, for all n,

$$f^{(2n)}(x) = AB^{n-1} + B^n f(x) \tag{45}$$

and, if $B \neq 0$,

$$\begin{aligned} \theta = E\{f[\xi(\mu)]\} &= f(\mu) + \sum_{n=1}^{\infty} \frac{1}{n!} [AB^{n-1} + B^n f(\mu)](\sigma^2/2)^n \\ &= f(\mu)e^{B\sigma^2/2} + \frac{A}{B} (e^{B\sigma^2/2} - 1). \end{aligned} \tag{46}$$

Alternatively, if $B = 0$, that is, if f is quadratic,

$$\theta = f(\mu) + A\sigma^2/2. \tag{47}$$

Similarly, for $B \neq 0$,

$$E[f(\hat{\mu})] = f(\mu)e^{B\lambda^2\sigma^2/2} + (A/B)(e^{B\lambda^2\sigma^2/2} - 1) \tag{48}$$

and, for $B = 0$,

$$E[f(\hat{\mu})] = f(\mu) + A\lambda^2\sigma^2/2. \tag{49}$$

Eliminating $f(\mu)$ from (46) and (48) and from (47) and (49), we find

$$\theta = e^{B(1-\lambda^2)\sigma^2/2}E[f(\hat{\mu})] + (A/B)[e^{B(1-\lambda^2)\sigma^2/2} - 1] \qquad \text{for } B \neq 0 \tag{50}$$

and

$$\theta = E[f(\hat{\mu})] + A(1 - \lambda^2)\sigma^2/2 \qquad \text{for } B = 0. \tag{51}$$

The last formula indicates that, when $B = 0$, the minimum variance unbiased estimate of θ is given by

$$\hat{\theta}(\hat{\mu}, S^2) = f(\hat{\mu}) + A(1 - \lambda^2)S^2/2\nu. \tag{52}$$

If $B \neq 0$ then, in order to obtain $\hat{\theta}(\hat{\mu}, S^2)$, it is sufficient to determine a function, say $\Phi(aS^2, \nu)$, independent of $\hat{\mu}$, such that its expectation equals $\exp(a\sigma^2/2)$. Taking into account the expansion

$$e^{a\sigma^2/2} = \sum_{n=0}^{\infty} \frac{1}{n!} (a\sigma^2/2)^n, \tag{53}$$

we easily find

$$\begin{aligned}\Phi(aS^2, \nu) &= \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\Gamma(\frac{1}{2}\nu)}{\Gamma(\frac{1}{2}\nu + n)} (aS^2/4)^n, \\ &= \left(\frac{2}{S\sqrt{a}}\right)^{\frac{1}{2}\nu-1} \Gamma(\tfrac{1}{2}\nu) I_{\frac{1}{2}\nu-1}(S\sqrt{a})\end{aligned} \tag{54}$$

where $I_n(x)$ is the Bessel function of imaginary argument.

It follows from (50) that

$$\hat{\theta}(\hat{\mu}, S^2) = \Phi[B(1 - \lambda^2)S^2, \nu][f(\hat{\mu}) + (A/B)] - (A/B), \tag{55}$$

which is the general formula for the minimum variance unbiased estimate of θ corresponding to the case where f is a recursive function. It will be seen that, generally, $\hat{\theta}(\hat{\mu}, S^2)$ is a linear function of the traditional estimate $f(\hat{\mu})$ of θ, with coefficients depending upon S^2, λ^2 and ν. If $\lambda^2 = 1$, that is, if the variance of $\hat{\mu}$ coincides with that of $\xi(\mu)$, then $\hat{\theta}(\hat{\mu}, S^2) = f(\hat{\mu})$. Otherwise $f(\hat{\mu})$ is biased. In the particular case $B = 0$, the correction for bias is additive, as indicated in (52). This makes the square root transformation very convenient (provided, of course, it provides effective normalization!) in dealing with balanced experiments in which the quantities to be estimated are differences of certain averages, the estimates of which all have the same variances.

If $A = 0$ but $B \neq 0$, then the correction for bias in $f(\hat{\mu})$ is multiplicative. Finally, if both A and B differ from zero, we have a combination of a multiplicative and an additive correction.

The importance of bias in the traditional estimate of θ may be evaluated by solving for $E[f(\hat{\mu})]$ equations (50) and (51). We have

$$E[f(\hat{\mu})] = [\theta + (A/B)]e^{-B(1-\lambda^2)\sigma^2/2} - (A/B), \qquad \text{for } B \neq 0 \tag{56}$$

and

$$E[f(\hat{\mu})] = \theta - A(1 - \lambda^2)\sigma^2/2, \qquad \text{for } B = 0. \tag{57}$$

10. Some particular cases. In this section we use the general results of Section 9 to deduce particular formulas. We obtain the minimum variance unbiased estimate of θ and the expectation of $f(\hat{\mu})$ referring to four particular normalizing transformations: (i) the square root transformation, (ii) the logarithmic transformation, (iii) the angular transformation and (iv) the hyperbolic sine transformation.

(i) In the case of the square root transformation, the transformed variable

$$\xi = (X - a)^{\frac{1}{2}} \tag{58}$$

where a is a known constant. We ignore the ambiguity connected with the fact that, for ξ to be a normal variable it must be capable of assuming negative values. The function f is

$$X = f(\xi) = \xi^2 + a. \tag{59}$$

Obviously this is a recursive function with $A = 2, B = 0$. Consequently, formula (52) yields directly

$$\hat{\theta}(\hat{\mu}, S^2) = f(\hat{\mu}) + (1 - \lambda^2)S^2/\nu = \hat{\mu}^2 + a + (1 - \lambda^2)S^2/\nu. \tag{60}$$

The bias of the traditional estimate $f(\hat{\mu}) = \hat{\mu}^2 + a$ is obtained from (57), namely,

$$E[f(\hat{\mu})] = \theta - (1 - \lambda^2)\sigma^2. \tag{61}$$

Hence, unless $\lambda^2 \geqq 1$, so that the variance of $\hat{\mu}$ is at least equal to that of $\xi(\mu)$, the use of $f(\mu)$ as an estimate systematically underestimates θ. Furthermore, the better the estimate $\hat{\mu}$, that is, the smaller the value of λ^2, the greater the bias.

(ii) In the case of logarithmic transformation, we have

$$\xi = \log_{10} X \tag{62}$$

and hence

$$X = f(\xi) = 10^\xi = e^{m\xi}, \quad \text{say}. \tag{63}$$

Here again the function f is of recursive type with $A = 0$ and $B = m^2$. Formula (55) gives

$$\begin{aligned} \hat{\theta}(\hat{\mu}, S^2) &= \Phi[m^2(1 - \lambda^2)S^2, \nu]f(\hat{\mu}) \\ &= \Phi[m^2(1 - \lambda^2)S^2, \nu]10^{\hat{\mu}}. \end{aligned} \tag{64}$$

Substituting $A = 0$ and $B = m^2$ into (56) we obtain

$$E[f(\hat{\mu})] = E10^{\hat{\mu}} = \theta e^{-m^2(1-\lambda^2)\sigma^2/2}. \tag{65}$$

Thus, with the logarithmic transformation, the bias of the traditional estimate is multiplicative. If the variance of $\hat{\mu}$ is less than σ^2 then the use of $10^{\hat{\mu}}$ will systematically underestimate θ and vice versa. The bias grows with increasing $|1 - \lambda^2|$.

(iii) With angular transformation we have

$$\xi = \arcsin \sqrt{X} \tag{66}$$

and

$$X = f(\xi) = \sin^2 \xi = \tfrac{1}{2}(1 - \cos 2\xi). \tag{67}$$

Here again the function f is of recursive type with $A = 2$ and $B = -4$. Hence, Formula (55) gives

$$\hat{\theta}(\hat{\mu}, S^2) = \Phi[4(\lambda^2 - 1)S^2, \nu]\,(\sin^2 \hat{\mu} - \tfrac{1}{2}) + \tfrac{1}{2}. \tag{68}$$

Substituting $A = 2$ and $B = -4$ in (56) we have

$$E[f(\hat{\mu})] = E\,[\sin^2 \hat{\mu}] = (\theta - \tfrac{1}{2})e^{2(1-\lambda^2)\sigma^2} + \tfrac{1}{2}. \tag{69}$$

It is seen that, if the variance of $\hat{\mu}$ is less than σ^2, the traditional estimate $\sin^2 \hat{\mu}$ is systematically "too far" from $\frac{1}{2}$. If the true value of $\theta < \frac{1}{2}$, then $\sin^2 \hat{\mu}$ will tend to underestimate θ. Otherwise, if $\theta > \frac{1}{2}$, there will be a tendency to overestimate θ. With $\lambda > 1$ these two tendencies will be reversed.

(iv) The last transformation to be considered here is based on the function

$$X = f(\xi) = \sinh^2 \xi = \tfrac{1}{2}\,[\cosh 2\xi - 1]. \tag{70}$$

It is of recursive type with $A = 2$ and $B = 4$. Hence, from Formula (55)

$$\hat{\theta}(\hat{\mu}, S^2) = \Phi[4(1 - \lambda^2)S^2, \nu]\,(\sinh^2 \hat{\mu} + \tfrac{1}{2}) - \tfrac{1}{2}. \tag{71}$$

Formula (57) with the indicated values of A and B gives

$$E[f(\hat{\mu})] = E\,[\sinh^2 \hat{\mu}] = (\theta + \tfrac{1}{2})e^{-2(1-\lambda^2)\sigma^2} - \tfrac{1}{2}. \tag{72}$$

In this case $f(\hat{\mu})$ underestimates or overestimates θ according to whether λ is less or greater than unity.

11. Concluding remarks. (i) Formula (54) defining Φ may seem complicated. In fact, the series on the right converges fairly rapidly so that sufficient accuracy is obtained with only a few terms.

It is easy to check that Formula (54) may be rewritten as follows

$$\Phi(aS^2, \nu) = \sum_{n=0}^{\infty} \frac{(aS^2/2)^n}{n! \prod_{k=1}^{n} (\nu + 2k - 2)} = \sum_{n=0}^{\infty} \frac{(a\hat{\sigma}^2/2)^n}{n! \prod_{k=1}^{n} \left(1 + \frac{(2k-2)}{\nu}\right)} \tag{73}$$

where

$$\hat{\sigma}^2 = S^2/\nu \tag{74}$$

is the unbiased estimate of σ^2. It will be seen that, for $n > 1$, the absolute value of each term on the right is less than the corresponding term in the expansion of $\exp\{a\hat{\sigma}^2/2\}$. In other words, the series (73) converges faster than the series for the exponential function.

(ii) In some circumstances, the practical statistician may decide to work on the assumption that the vairance σ^2 is known. In order to adjust the formulas deduced in this paper to this case, it is sufficient to replace $\hat{\sigma}^2$ by σ^2 and pass to the limit as $\nu \to \infty$. In particular, this procedure reduces the right hand side of (73) to $\exp\{a\sigma^2/2\}$, which is (53).

(iii) Formulas have been published for correcting the bias introduced by the transformation of variables in some particular cases in [1], [6] and [7], for example. However, these formulas do not agree with ours.

(iv) It is a pleasure to express our indebtedness to the referee who picked up several mistakes in the original text of the paper and, in addition, called our

attention to the important publications [3], [4] and [9], which we overlooked. Among the problems treated in these papers, there is one which is strongly related to ours. In the present notation, this problem consists in finding the function $h(\hat{\mu}, S^2, \nu)$ that is an unbiased estimate of a given function $g(\mu, \sigma^2)$. This problem is treated under the restriction that $\lambda^2 = 1/(\nu + 1)$. Apart from this restriction, in order to reduce our problem to the problem just described, it is sufficient to evaluate the expectation $E\{f[\xi(\mu)]\}$ and to denote the result by $g(\mu, \sigma^2)$. The difference between our results and those in [3], [4] and [9] consists first in a difference in the method and in the conditions of the various theorems: in the earlier papers the conditions are expressed in terms of the function g whereas, in the present paper, they refer to the function f. Also, explicit formulas for the unbiased estimates of θ, as given here, are not contained in the papers quoted.

REFERENCES

[1] F. J. Anscombe, "The transformation of Poisson, binomial, and negative-binomial data," *Biometrika*, Vol. 35 (1948), pp. 246–254.
[2] M. S. Bartlett, "The use of transformations," *Biometrics*, Vol. 3 (1947), pp. 39–52.
[3] Tosio Kitagawa, "The operational calculus and the estimations of functions of parameter admitting sufficient statistics," *Bull. Math. Stat.*, Vol. 6 (1956), pp. 95–108.
[4] A. N. Kolmogoroff, "Unbiased estimates," *Izvestiya Akademii Nauk SSSR, Ser. Mat.*, Vol. 14 (1950), pp. 303–326; see also *Am. Math. Soc. Translation* No. 98.
[5] E. L. Lehmann and H. Scheffé, "Completeness, similar regions and unbiased estimation, Part I," *Sankhyā*, Vol. 10 (1950), pp. 305–340.
[6] H. A. Meyer, "A correction for systematic error occurring in the application of the logarithmic volume equation," *J. Series Penn. Agric. Exp. Stat.*, Paper 1058 (1941), pp. 1–3.
[7] M. H. Quenouille, *The Design and Analysis of Experiment*, Hafner, New York, 1953.
[8] John W. Tukey, "On the comparative anatomy of transformations," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 602–632.
[9] Yasutoshi Washio, Haruki Morimoto, and Nobuyuki Ikeda, "Unbiased estimation based on sufficient statistics," *Bull. Math. Stat.*, Vol. 6 (1956), pp. 69–93.

ON A PROBLEM OF J. NEYMAN AND E. SCOTT[1]

By L. Schmetterer

University of Hamburg and University of California, Berkeley

1. Introduction. Let ξ be a one-dimensional random variable distributed according to $N(a, \sigma^2)$ (that means a normal distribution with mean value a and variance σ^2) and $f(x)$ a measurable function defined on $-\infty < x < \infty$. Suppose that

$$((2\pi)^{\frac{1}{2}}\sigma)^{-1} \int_{-\infty}^{+\infty} f(x)e^{-(x-a)^2/2\sigma^2}\,dx = \omega(a, \sigma^2)$$

exists for all real numbers a and all $\sigma^2 > 0$ as a Lebesgue integral. Let λ be a fixed positive number and η a one-dimensional random variable with distribution $N(a, \sigma^2\lambda^2)$. Let ζ be a random variable such that z/σ^2 has a Pearson-Helmert distribution with ν degrees of freedom. Further, we suppose that η and ζ are independent. The question is whether or not there are unbiased estimates $H(\eta, \zeta)$ for $\omega(a, \sigma^2)$ where $-\infty < a < \infty$, and where $\sigma^2 > 0$. In a paper of Neyman and Scott [1] it is proved that there is always an unbiased estimate $H(\eta, \zeta)$ for $\omega(a, \sigma^2)$ for the class of entire functions $f(x)$ which satisfy the conditions

$$(1) \qquad \frac{1}{n}(|f^{(2n)}(0)|)^{1/n} = o(1), \qquad \frac{1}{n}(|f^{(2n+1)}(0)|^{1/n} = o(1)$$

and which take real values on the real line. Condition (1) can be expressed in the following way: $f(x)$ is an entire function of order 2 and type zero or of any smaller order. Let us recall that an entire function $f(x)$ is of order k and type $\alpha \geqq 0$ if $|f(re^{i\varphi})| = O(\exp\{(\alpha + \epsilon)r^k\})$ for every $\epsilon > 0$ but for no $\epsilon < 0$ if $r \geqq r(\epsilon)$ and $0 \leqq \varphi < 2\pi$.

In addition, the following problem is raised in the paper just mentioned: Let $f(x)$ be a measurable function defined on the whole real line such that

(i) $\int_{-\infty}^{+\infty} |f(x)|\, e^{-\epsilon x^2}\,dx$ converges for all $\epsilon > 0$.

Is there always an unbiased estimate $H(\eta, \zeta)$ for $\omega(a, \sigma^2)$ if $f(x)$ satisfies Condition (i)? In this paper it will be shown that there is for each $f(x)$ satisfying Condition (i) an unbiased estimate $H(\eta, \zeta)$ for $\omega(a, \sigma^2)$ if λ is any positive number in the interval $0 < \lambda \leqq 1$. However, if λ is allowed to be >1, then the unbiased estimate $H(\eta, \zeta)$ need not exist. Also, it will be shown that the theorem of Neyman and Scott mentioned above is, in a certain sense, best.

Received October 2, 1959; revised February 6, 1960.

[1] This paper was written while the author was a research professor at the Adolph C. and Mary Sprague Miller Institute for Basic Research in Science.

2. Problem and solution. Let us now consider for any $\lambda > 0$ and any real number $\nu \geqq 1$ the integral equation

$$\begin{aligned}(2)\qquad &\int_{-\infty}^{+\infty} f(x)e^{-(x-a)^2/2\sigma^2}\,dx \\ &= (2\lambda\sigma^2\Gamma(\nu/2))^{-1}\int_{-\infty}^{+\infty}\int_0^{\infty} H(y,z)e^{-(y-a)^2/2\sigma^2\lambda^2}e^{-z/2\sigma^2}(z/2\sigma^2)^{(\nu/2)-1}\,dy\,dz,\end{aligned}$$

where $-\infty < a < \infty$, $\sigma^2 > 0$ and $f(x)$ is any measurable function satisfying (i). We will prove the following theorem.

THEOREM 1. *If $f(x)$ is a measurable function defined on the whole real line and satisfying* (i), *then for all positive numbers λ in the interval $0 < \lambda \leqq 1$ there is a solution $H(y, z)$ of* (2) *which satisfies the following "natural" condition*

(ii) $\int_{-\infty}^{+\infty}\int_0^{\infty} |H(y,z)|\, z^{(\nu/2)-1}e^{-\epsilon y^2}e^{-\eta z}\,dy\,dz$ *converges for all $\epsilon > 0$, $\eta > 0$.*

Moreover, there is only one solution of this kind (up to sets of measure zero, of course). For $0 < \lambda < 1$ and $\nu > 1$ this solution is given by

$$\begin{aligned}H(y,z) = \Gamma\left(\frac{\nu}{2}\right)[(\pi)^{1/2}(1-\lambda^2)^{1/2}\Gamma((\nu-1)/2)]^{-1}z^{-(\nu/2)+1} \\ \cdot\int_{-\infty}^{+\infty} f(x)[I(z-(x-y)^2/(1-\lambda^2))]^{(\nu-3)/2}\,dx\end{aligned}$$

$$-\infty < y < \infty,\quad 0 < z < \infty,$$

where

$$(I(x))^{\alpha} = \begin{cases} x^{\alpha} & x > 0 \\ 0 & x \leqq 0.\end{cases}$$

for every real number α. For $\nu = 1$ the solution is given by

$$H(y,z) = \tfrac{1}{2}\{f[y-(z(1-\lambda^2))^{1/2}] + f[y+(z(1-\lambda^2))^{1/2}]\}.$$

For $\lambda = 1$ the solution is given by

$$H(y,z) = f(y),\qquad -\infty < y < \infty,\quad 0 \leqq z < \infty.$$

For the proof we observe first that the assertion concerning $\lambda = 1$ is trivial. For $\lambda < 1$, notice that $f(x)$ is locally integrable by (i). Therefore, it is obvious that

$$J(y,z,\nu) = \int_{-\infty}^{+\infty} f(x)\left(I\left(z - \frac{(x-y)^2}{1-\lambda^2}\right)\right)^{(\nu-3)/2} dx$$

is defined for all y, all nonnegative z and for $\nu \geqq 3$. Moreover, it is easy to show that $J(y, z, \nu)$ defines, for all y, each real $\nu > 1$ and almost all nonnegative z, a locally integrable function of z, which after multiplication by $\exp\{-\eta z\}$ for any $\eta > 0$ is absolutely integrable for $0 < z < \infty$. To show this, consider the

identity

$$
(3)\quad \begin{aligned} \Gamma\left(\frac{\nu-1}{2}\right)(2\sigma^2)^{(\nu-1)/2} &\int_{-\infty}^{+\infty} f(y+t)e^{-t^2/2\sigma^2(1-\lambda^2)}\,dt \\ &= \int_0^\infty u^{(\nu-3)/2}e^{-u/2\sigma^2}\,du \int_{-\infty}^{+\infty} f(y+t)e^{-t^2/2\sigma^2(1-\lambda^2)}\,dt. \end{aligned}
$$

The iterated integral on the right side of (3) is absolutely convergent by (i) for each positive $\lambda < 1$, each $\sigma^2 > 0$, each $\nu > 1$ and all real y. But easy transformation of the integration variables shows that this iterated integral is equal to

$$
(1-\lambda^2)^{\frac{1}{2}}2^{-1}\int_0^\infty u^{(\nu-3)/2}\int_u^\infty e^{-z/2\sigma^2}(f(X)+f(Y))(z-u)^{-\frac{1}{2}}\,dz\,du
$$

where $X = y + [(z-u)(1-\lambda^2)]^{\frac{1}{2}}$ and where $Y = y - [(z-u)(1-\lambda^2)]^{\frac{1}{2}}$. The absolute convergence of this iterated integral justifies changing the order of integration and after an easy calculation we get for this integral

$$
\int_0^\infty e^{-z/2\sigma^2}\int_{-\infty}^{+\infty} f(x)\left(I\left(z-\frac{(x-y)^2}{1-\lambda^2}\right)\right)^{(\nu-3)/2}\,dx\,dz.
$$

Another application of Fubini's theorem and the fact that $\exp\{-z/2\sigma^2\}$ is positive and bounded for all nonnegative z and every $\sigma^2 > 0$ show that $J(y, z, \nu)$ is a locally integrable function of z for all y, each $\nu > 1$, and almost all nonnegative z. This function is, after multiplication by $\exp\{-\eta z\}$, for any $\eta > 0$ absolutely integrable for all nonnegative z. Now we use the following simple lemma.

LEMMA. *If $f(x)$ satisfies* (i) *then for any λ in $0 < \lambda < 1$ there exists exactly one solution $g(y, \sigma^2)$ of the equation*

$$
(4)\quad \int_{-\infty}^{+\infty} f(x)e^{-(x-a)^2/2\sigma^2}\,dx = \frac{1}{\lambda}\int_{-\infty}^{+\infty} g(y,\sigma^2)e^{-(y-a)^2/2\sigma^2\lambda^2}\,dy, \qquad -\infty < a < \infty,
$$

for which the integral on the right side of (4) *converges absolutely for each $\sigma^2 > 0$ and all real a. This solution is given by*

$$
(5)\quad g(y,\sigma^2) = [(2\pi)^{\frac{1}{2}}\sigma(1-\lambda^2)^{\frac{1}{2}}]^{-1}\int_{-\infty}^{+\infty} f(x)e^{-(x-y)^2/2\sigma^2(1-\lambda^2)}\,dx, \qquad -\infty < y < \infty.
$$

The uniqueness for a solution of (4) with the asserted property is well known. For the proof that (5) satisfies (4) it is enough to notice that

$$
\begin{aligned} [(2\pi)^{\frac{1}{2}}\sigma((1-\lambda^2)^{\frac{1}{2}})^{-1}] \int_{-\infty}^{+\infty} f(x)\int_{-\infty}^{+\infty} &e^{-(y-x)^2/2\sigma^2(1-\lambda^2)}e^{-(y-a)^2/2\sigma^2\lambda^2}\,dy\,dx \\ &= \lambda\int_{-\infty}^{+\infty} f(x)e^{-(x-a)^2/2\sigma^2}\,dx \end{aligned}
$$

and that the iterated integral on the left side of this identity converges absolutely by (i).

Now we proceed with the proof of Theorem 1. Using (3) and the above lemma it follows (again by an application of Fubini's theorem) that $J(y, z, \nu)$ is a locally integrable function for almost all y, $-\infty < y < \infty$, and almost all z, $0 < z < \infty$, and that $H(y, z)$ satisfies the integral equation (2). The lemma just mentioned and the remarks about the absolute integrability of $\exp\{-\eta z\} J(y, z, \nu)$ for all $\eta > 0$ show that $H(y, z)$ satisfies condition (ii). The asserted representation of $H(y, z)$ for the case $\nu = 1$ is now obvious. Moreover, it is well known that there is only one solution $H(y, z)$ of (1) satisfying (ii).

3. A negative result. Now we will prove that in a certain sense the result obtained by Neyman and Scott concerning the existence of unbiased estimates $H(\eta, \zeta)$ for $\omega(a, \sigma^2)$ cannot be improved.

THEOREM 2. *Let $f(x)$ be an entire function with real values on the real line and satisfying Condition* (i). *There is for each real $\alpha > 0$ an entire function of order 2 and type α such that* (2) *does not have a solution $H(y, z)$ which satisfies* (ii) *for any $\lambda > 1$. This means that there is no unbiased estimate for $\omega(a, \sigma^2)$ according to the usual definition of unbiased estimates.*

For the proof take $f(x) = \exp\{-\alpha x^2\}$. For a given $\alpha > 0$, the function $f(x)$ is an entire function of order 2 and type α, which takes real values on the real line. Suppose there is a solution $H(y, z)$ of (2) which satisfies Condition (ii). Then for each $\sigma^2 > 0$, $K(y, \sigma^2)$ defined by

$$K(y, \sigma^2) = (\Gamma(\nu/2))^{-1} \int_0^\infty H(y, z) e^{-z/2\sigma^2} (z/2\sigma^2)^{(\nu/2)-1} \, dz/2\sigma^2$$

exists for almost all y, is locally integrable, and

$$\int_{-\infty}^{+\infty} K(y, \sigma^2) e^{-(y-a)^2/2\sigma^2\lambda^2} \, dy$$

exists as an absolutely convergent integral for all real a, each $\sigma^2 > 0$ and any $\lambda > 0$. Hence, for each $\lambda > 1$ and each $\sigma^2 > 0$ the equation

$$\int_{-\infty}^{+\infty} e^{-\alpha x^2} e^{-(x-a)^2/2\sigma^2} \, dx = 1/\lambda \int_{-\infty}^{+\infty} K(y, \sigma^2) e^{-(y-a)^2/2\sigma^2\lambda^2} \, dy$$

must be an identity in a for all real a. We write $a = -s$ and get

$$(6) \qquad \left[\frac{2\pi\sigma^2}{2\alpha\sigma^2 + 1}\right]^{\frac{1}{2}} e^{-\alpha s^2/(2\alpha\sigma^2+1)} e^{s^2/2\sigma^2\lambda^2} = \frac{1}{\lambda} \int_{-\infty}^{+\infty} K(y, \sigma^2) e^{-y^2/2\sigma^2\lambda^2} e^{-ys/\sigma^2\lambda^2} \, dy.$$

But it is obvious that these two expressions define for each $\sigma^2 > 0$ and each $\lambda > 1$ analytical functions of s in the whole complex plane, which are identical on the real line and so, everywhere. Take now

$$(7) \qquad \sigma^2(\lambda^2 - 1) > 1/2\alpha.$$

This is possible since $\lambda > 1$. For $s = it$ and real t the right side of (6) is a Fourier transform of an absolutely integrable function and must converge to zero for $|t| \to \infty$. But the left side of (6) goes to infinity for $|t| \to \infty$ by (7). This is a contradiction to the existence of an unbiased estimate for $\omega(a, \sigma^2)$.

Remark 1. We notice that the same argument proves the nonexistence of more general solutions $H(y, z)$ of (2) in the case $\lambda > 1$. For instance, it is easy to show that for $f(x) = \exp\{-\alpha x^2\}$, $\alpha > 0$, there is no solution $H(y, z)$ of (2) for which the integral on the right side of (2) can be written as an iterated conditionally convergent integral of the form

$$\left[\lambda\Gamma\left(\frac{\nu}{2}\right)\right]^{-1}(2\sigma^2)^{-\nu/2}\lim_{M,N\to\infty}\int_{-M}^{N} e^{-(y-a)^2/2\sigma^2\lambda^2}\lim_{\substack{\epsilon\to 0\\ P\to\infty}}\int_{\epsilon}^{P} H(y, z)e^{-z/2\sigma^2}z^{\nu/2-1}\,dz\,dy.$$

For this we have only to consider the identity in s obtained from (6) by modifying the definition of $K(y, \sigma^2)$ and the right side of (6) in an obvious manner. But using (allowed) partial integration it is easy to show that now the integral on the right side of (6) exists also for all s of the complex plane. Further, it is well known that the Fourier transform of a conditionally integrable function is a $o(|t|)$ for $|t| \to \infty$ and this leads again to a contradiction if σ^2 satisfies (7).

Remark 2. Using Laplace transforms it is also possible to give a more general theorem than Theorem 1 which covers also the case $\lambda > 1$. Because Theorem 2 gives a clear idea in which way such a theorem must be formulated, I do not think that it is of any value to give it here.

4. Examples. According to Theorem 1 there is a solution $H(y, z)$ of (2) for any λ in the interval $0 < \lambda \leqq 1$ if $f(x)$ is given by $\exp\{-\alpha x^2\}$, $\alpha > 0$. For $0 < \lambda < 1$ this solution is given by

$$Ce^{-\alpha y^2}z^{-(\nu/2)+1}\int_0^z e^{-\alpha u(1-\lambda)^2}(z-u)^{(\nu-3)/2}u^{-\frac{1}{2}}\cosh\left(2\alpha y(u(1-\lambda^2))^{\frac{1}{2}}\right)du,$$

where $C = \Gamma(\nu/2)[\Gamma((\nu-1)/2)\sqrt{\pi}]^{-1}$. This expression and Theorem 2 give a complete answer to a question raised in the paper of Neyman and Scott (p. 12 of [1]).

Another function $f(x)$ whose inverse is occasionally used as a normalizing transform is given by

$$f(x) = \begin{cases} (b+x)^p, & p > 0 \qquad x > -b \\ 0 & \qquad\qquad x \leqq -b. \end{cases}$$

If p is an integer it is more usual to define $f(x)$ by $(b + x)^p$ for all real x. Moreover in most practical applications $|b|$ is a very large number such that these two definitions are "almost" equivalent. It appears that there is an unbiased estimate $H(\eta, \zeta)$ in the case $0 < \lambda < 1$, but it does not seem to be appropriate for practical use in general, although it is of course possible for rational numbers p and integers $\nu = 2k + 3$, where $k \geqq 0$, to express $J(y, z, \nu)$ by elementary functions. For instance, for $y > -b$ and $z < (b + y)^2/(1 - \lambda^2)$ we get the following form of $H(y, z)$ where $p = n/m$, with $(n, m) = 1$.

$$H(y, z) = m\Gamma(\nu/2)[\sqrt{\pi}((1 - \lambda^2)^{\frac{1}{2}})\Gamma((\nu - 1)/2)]^{-1}z^{-(\nu/2)+1} \sum_{l=0}^{k} z^l \binom{k}{l} (-1)^{k-l}$$

$$\cdot (1 - \lambda^2)^{-(k-l)} \left[X^{(m+n)/m} \sum_{r=0}^{2(k-l)} (-1)^r \binom{2(k-l)}{r} \frac{(a + y)^r X^{2(k-l)-r}}{(n + m)[1 + 2(k - l) - r]} \right.$$

$$\left. - Y^{(m+n)/m} \sum_{r=0}^{k} (-1)^r \binom{2(k-l)}{r} \frac{(a + y)^r Y^{2(k-l)-r}}{(n + m)[1 + 2(k - l) - r]} \right]$$

where $X = b + y + [z(1 - \lambda^2)]^{\frac{1}{2}}$ and where $Y = b + y - [z(1 - \lambda^2)]^{\frac{1}{2}}$. If p is an integer and $f(x)$ is defined by $(b + x)^p$ for all x then $H(y, z)$ is given by this expression for all y and $z \geqq 0$ and any $\lambda > 0$ in agreement with the results of Neyman and Scott.

5. Relation to previously obtained results. The proof of Theorem 1 shows that

$$h_x(y, z) = C(\lambda, \nu)z^{-(\nu/2)+1}[I(z - (x - y)^2/(1 - \lambda^2))]^{(\nu-3)/2},$$
$$-\infty < y < \infty, \quad 0 < z < \infty$$

where

$$C(\lambda, \nu) = \Gamma\left(\frac{\nu}{2}\right)[\sqrt{\pi}(1 - \lambda^2)^{\frac{1}{2}}\Gamma((\nu - 1)/2)]^{-1}$$

is a solution of the equation

$$e^{-(x-a)^2/2\sigma^2} = (2\lambda\sigma^2\Gamma(\nu/2))^{-1} \int_{-\infty}^{+\infty}\int_0^{\infty} h_x(y, z)e^{-(y-a)^2/(2\sigma^2\lambda^2)}e^{-z/2\sigma^2} \cdot (z/2\sigma^2)^{(\nu/2)-1}\, dy\, dz$$

for every $\nu > 1$, every fixed real x and $0 < \lambda < 1$. This means $h_x(\eta, \zeta)$ is an unbiased estimate for $e^{-(x-a)^2/2\sigma^2}$. This fact is proved and used by Kolmogorov [2], who obtains the results of Theorem 1 for the following special cases: $f(x)$ is the characteristic function of a measurable set and $\lambda = (1/(\nu + 1))^{\frac{1}{2}}$, $\nu = 2, 3, 4, \cdots$. But at least for integers $\nu \geqq 2$ it is easy, using Kolmogorov's method, to extend his result to the more general assertion of Theorem 1. The method of Kolmogorov and my own method for proving Theorem 1 are of course closely related. Another paper which can be mentioned in connection with the method used here is one by Washio, Morimoto and Ikeda [3].

6. Acknowledgments. I am grateful to Mr. Neyman and Miss Scott for the privilege of reading their paper before its publication. Further I am indebted to the referee for valuable comments. Especially the content of Section 5 arose out of one of his suggestions.

REFERENCES

[1] J. Neyman and E. L. Scott, "Correction for bias introduced by a transformation of variables," *Ann. Math. Stat.*, Vol. 31 (1960) pp. 643–655.

[2] A. N. Kolmogorov, "Unbiased estimates", *Izvestiya Akademii Nauk SSSR*. Seriya Matematiceskaya, Vol. 14 (1950), pp. 303–326. Amer. Math. Soc. Trans. No. 98.

[3] Y. Washio, H. Morimoto, N. Ikeda, "Unbiased estimation based on sufficient statistics," *Bull. Math. Stat.*, Vol. 6 (1956), pp. 69–93.

ESTIMATES WITH PRESCRIBED VARIANCE BASED ON TWO-STAGE SAMPLING

By Allan Birnbaum[1] and William C. Healy, Jr.[2]

New York University and Ethyl Corporation

1. Summary. A method is given which provides, under conditions satisfied by many common distributions, rules for sampling in two stages so as to obtain an unbiased estimator of a given parameter, having variance equal to, or not exceeding, a prescribed bound. The method is applied to estimation of the means of binomial, Poisson, and hypergeometric distributions; scale-parameters in general and of the Gamma distribution in particular; the variance of a normal distribution; and a component of variance. The use of such estimators to achieve homoscedasticity is discussed. Optimum sampling rules are discussed for some of these estimators, and some tables are given to facilitate their use. The efficiency of the method is shown to be high in many cases.

2. Introduction. In most problems of estimation, estimators based on samples of fixed sizes have precisions which depend on unknown parameters, and estimators with prescribed precision are not available without resort to sequential sampling in two or more stages, as in Stein's procedure [1] for estimation of the mean of a normal distribution with unknown variance. For problems other than those of the type treated by Stein the only available general methods which are both fairly practicable and efficient seem to be the double-sampling method of Cox [2], [3] and the sequential method of Anscombe [4]. The latter methods, however, are approximate, being based on asymptotic theory, and there seems to be no easily applicable method available for determining in a given case the closeness of the approximations involved. An approach employing a different concept of prescribed precision is described by Graybill [16].

The method to be described below (developed independently by the authors) is a simple one which provides, in a number of problems, procedures for two-stage sampling leading to estimators which are exactly unbiased; in certain problems these estimators have exactly a prescribed variance, while in other problems they have variances never exceeding but generally close to a prescribed bound. Under certain conditions, primarily that the precision prescribed is sufficiently high, these estimators are shown to have generally high efficiency.

3. General discussion of the method.

3.1 Statement of problems. Let $S = \{x\}$ be the sample space for a single random observation, X, on which a density or discrete elementary probability function,

---

Received February 12, 1959; revised October 16, 1959.

[1] Work supported by the Office of Naval Research.

[2] Work supported in part by the Office of Ordinance Research, U. S. Army, while this author was at the University of Illinois.

$f(x, \theta)$, is defined for each θ in a given parameter space Ω. Suppose that it is desired to estimate with prescribed precision a real-valued function $\rho = \rho(\theta)$. We adopt the following formalization of this requirement: it is required to find find an unbiased estimator of ρ having variance not exceeding a given positive function $B(\theta)$.

3.2 Assumptions.

I. Assume that, for each non-sequential sample size n not less than a known n_0, there exists an unbiased estimator $t = t(x_1, \cdots x_n)$ of ρ, i.e.,

$$E_\theta t(X_1, \cdots X_n) = \rho(\theta). \tag{1}$$

II. Let $\sigma^2(\theta, n)$ denote the variance of t for sample size n. Assume that, for each non-sequential sample size m not less than a known m_0, there exists a measurable "second sample size function" $n = n(x_1, \cdots x_m)$ taking integer values not less than n_0, and such that either

$$E_\theta \sigma^2(\theta, n(X_1, \cdots X_m)) = B(\theta) \tag{2a}$$

or

$$E_\theta \sigma^2(\theta, n(X_1, \cdots X_m)) \leqq B(\theta) \tag{2b}$$

holds for each θ.

3.3 Estimation Procedure. Under the above assumptions, a simple unbiased estimator of ρ, having variance not exceeding $B(\theta)$, is given by any procedure of the following form:

A) Take a sample of m observations ($m \geqq m_0$), $x_1, \cdots x_m$, and compute $n = n(x_1, \cdots x_m)$.

B) Take a second independent sample of $n = n(x_1, \cdots x_m) \geqq n_0$ additional observations, $x_{m+1}, \cdots x_{m+n}$.

C) Estimate ρ by $t = t(x_{m+1}, \cdots x_{m+n})$, ignoring at this stage the first sample observations $x_1, \cdots x_m$.

The fact that this procedure seems to involve gross waste of information in the first sample suggests at first sight that its efficiency must be low. It will be shown, however, that the efficiency of the method, with a suitable choice of sampling rule, is so high in a number of cases that the search for more efficient methods (generally not known at present) would seem to be of more theoretical than practical interest for those cases.

3.4 Properties of the Method. We first verify that, when functions t and n can be found satisfying conditions (1) and (2b) above, the method gives unbiased estimators with variances not exceeding the prescribed bound. Let

$$N = n(X_1, \cdots X_m).$$

Then the estimate is $T = t(X_{m+1}, \cdots X_{m+n})$, and

$$E_\theta(T) = E_N(E_T[T \mid N]) = E_\theta(\rho) = \rho, \tag{3}$$

since, for each fixed $n \geqq n_0$, we have $E_\theta t(X_{m+1}, \cdots X_{m+n}) = \rho$ by (1). Also

$$\text{Var}_\theta (T) = E_N \text{Var}_T (T \mid N) = E_\theta \sigma^2(\theta, N) \leqq B(\theta) \tag{4}$$

by (2b); if (2a) holds, $\text{Var}_\theta (T) = B(\theta)$.

3.5 Efficiency Considerations. A measure of efficiency for any sequential estimator satisfying (3) and (4), and *not* restricted to the use of only two stages of sampling, may be devised as follows: It has been shown by Wolfowitz [5], under certain regularity conditions on $f(x, \theta)$ and $\rho(\theta)$ and certain broad conditions on the sequential sampling rule, that each unbiased sequential estimator t of ρ, together with its total random sequential sample size N', satisfies

$$\text{Var}_\theta(T) \geqq E_\theta \left(\frac{\partial \log f(X, \theta)}{\partial \rho}\right)^2 \Big/ E_\theta(N'). \tag{5}$$

From (4) and (5) we obtain, under the conditions mentioned, the following lower bound for the expected total sample size required by any sequential estimator meeting our conditions (3) and (4):

$$E_\theta(N') \geqq E_\theta \left(\frac{\partial \log f(X, \theta)}{\partial \rho}\right)^2 \Big/ B(\theta) \tag{6}$$

As will be shown by specific examples below, there does not necessarily exist an estimator which attains this lower bound. Nevertheless it is useful to define as an index of efficiency the function

$$R(\theta) = E_\theta \left(\frac{\partial \log f(X, \theta)}{\partial \rho}\right)^2 \Big/ [B(\theta)E_\theta(N')], \tag{7}$$

where $E_\theta(N')$ is computed for any given estimate satisfying (3) and (4). As an example of the interpretation of this index, suppose that for a given estimator we find that $R(\theta) \geqq 0.90$ for all θ; then we can assert that for every estimator meeting the conditions (3), (4), and the general conditions of [5], the required expected total sample size function $E_\theta(N^*)$ will satisfy $E_\theta(N^*) \geqq 0.90\ E_\theta(N')$ for all θ; hence average savings in sample sizes of at most 10% might be achieved. It is known that in general the savings actually possible are less than indicated by such bounds, e.g. less than the 10% indicated here.) Such efficiency bounds are given in the following sections for various specific problems.

The estimation methods of the present paper are roughly similar to the method of Stein [1] for estimation of a normal mean. For most purposes the prescribed-length confidence interval formulation adopted by Stein seems preferable to the prescribed-variance formulation adopted here; the present formulation is akin to a decision-theoretic one with mean-squared error loss function, but the restriction of unbiasedness which provides essential simplications of calculations also generally entails some inefficiency from this standpoint. While Stein was able to give exact confidence intervals by determination of the exact (Student's) distribution of the point estimator implicit in his method, the exact distributions of the estimators given here are not known. Consequently this paper makes

no contribution to the theory of exact interval estimation comparable with Stein's, apart from the following crude use of Tchebycheff's inequality: If $\hat{\theta}$ is an unbiased estimator of θ with variance not exceeding a constant B, then the interval estimator $\hat{\theta} \pm \epsilon$ covers θ with probability at least $1 - \alpha$ where $\alpha = B/\epsilon^2$. For many of the problems considered below, even such confidence intervals have not previously been available. (For a number of problems, a method of constructing confidence intervals of fixed length and confidence coefficient, but probably poor efficiency, was given in [6]).

In many cases, particularly those in which high precision is specified, the estimators given here have approximately normal distributions. This is illustrated in the Poisson case below. To the extent that this is true, all methods for confidence regions and significance tests based on assumptions of normality with known variance may be applied. Useful approximations to the distributions of some of the estimators can probably be based on Student's t distributions with the number of degrees of freedom determined by a fitting of fourth moments; further investigation of this possibility is required.

It should be noted that, with the methods of this paper, there will sometimes occur samples which on inspection strongly suggest that some modification of the estimators given here would be more appropriate and efficient. A similar comment applies, with somewhat less force, to Stein's procedure and some other sequential procedures. These features seem symptomatic of possible improvements in efficiency of these methods which have not yet been found. They seem also to point to more basic problems in the foundations of statistical inference which lie outside the scope of the present paper. The estimators given in this paper have variance and efficiency properties which are valid within the unconditional two-stage sampling probability framework; these properties are not considered here (except in some computational steps) conditionally on a given first or second sample size. The unbiasedness properties of these estimators generally hold both conditionally and unconditionally.

4. Estimation of a mean. Suppose that X is real-valued and that the mean

$$\theta = \rho(\theta) = E_\theta(X)$$

is the parameter to be estimated. Then Assumption I is obviously satisfied if we take

$$t = t(x_{m+1}, \cdots x_{m+n}) = \frac{1}{n} \sum_{i=1}^{n} x_{m+i} .$$

Letting $\sigma_\theta^2 = \text{Var}_\theta (X)$, we have $\sigma^2(\theta, n) = \sigma_\theta^2/n$. Condition (4) becomes

$$E_\theta(1/N) = E_\theta 1/n(X_1, \cdots X_m) \leq B(\theta)/\sigma_\theta^2 . \tag{8}$$

Then any integer $m \geq m_0$, and any function $n = n(x_1, \cdots x_m)$ satisfying (8), may be used to define an estimator, which will then automatically satisfy (3) and (4). Such an estimator has expected total sample size

$$E_\theta(N') = m + E_\theta[n(X_1, \cdots X_m)] = m + E_\theta(N), \tag{9}$$

and efficiency bound

$$(10) \qquad R(\theta) = E_\theta\left(\frac{\partial \log f(X, \theta)}{\partial \theta}\right)^2 \Big/ B(\theta)\,[m + E_\theta(N)].$$

In the special case of constant prescribed precision, $B(\theta) \equiv B$, (8) becomes $(1/B)E_\theta[1/n(X_1, \cdots X_m)] \leqq 1/\sigma_\theta^2$, and the problem of finding a suitable second-sample-size function $n(x_1, \cdots x_m)$ may be stated as the problem of finding an estimator $1/\hat{\sigma}^2$ of $1/\sigma_\theta^2$, based on m observations, which is unbiased (condition (2a)), or which has positive bias at no $\theta \varepsilon \Omega$ (condition (2b)). Then the sequential sampling rule may be stated as:

A′) Observe $x_1, \cdots x_m$, and compute $\hat{\sigma}^2 = \hat{\sigma}^2(x_1, \cdots x_m)$.
B′) Take a second sample of $n = \hat{\sigma}^2/B$ observations $x_{m+1}, \cdots x_{m+n}$.
C′) Estimate the mean $\theta = E_\theta(X)$ by the mean $t = 1/n\sum_{i=1}^n x_{m+i}$ of the second sample only.

It is sometimes convenient to define $\hat{\sigma}^2(x_1, \cdots, x_m)$ formally in such a way that $\hat{\sigma}^2/B$ is not always an integer. Then for most applications it will suffice to take n as the smallest integer not less than $\hat{\sigma}^2/B$. A calculation like that above shows that this gives again $\text{Var}_\theta(T) \leqq B$. Alternatively, given $\hat{\sigma}^2/B$, we could use a random device to choose $n = [\hat{\sigma}^2/B]$ = the largest integer not exceeding $\hat{\sigma}^2/B$, with a probability γ, and $n = [\hat{\sigma}^2/B] + 1$ with probability $1 - \gamma$, where γ is determined by the equation

$$\gamma[\hat{\sigma}^2/B]^{-1} + (1 - \gamma)([\hat{\sigma}^2/B] + 1)^{-1} = B/\hat{\sigma}^2.$$

The latter procedure, which is perhaps of primarily theoretical interest, gives $\text{Var}_\theta(T) = B$ exactly if $E_\theta[1/\hat{\sigma}^2(X_1, \cdots X_m)] = 1/\sigma^2$ exactly. Henceforth we write $n = \hat{\sigma}^2/B$ to indicate that one of these procedures is used in defining n. It follows that calculations based on the equation $n = \hat{\sigma}^2/B$, such as the equation

$$E_\theta n(X_1, \cdots X_m) = E_\theta\hat{\sigma}^2(X_1, \cdots X_m)/B$$

used below, may involve an error whose magnitude is in any case less than one. Similar remarks apply to cases of the method other than those of estimation of a mean.

For any such procedure we have expected total sample size

$$E_\theta(N') = m + E_\theta n(X_1, \cdots X_m) = m + (1/B)E_\theta\hat{\sigma}^2(X_1, \cdots X_m),$$

and efficiency bound given by

$$1/R(\theta) = (B/\sigma_\theta^2)E_\theta(N') = (Bm/\sigma_\theta^2) + (1/\sigma_\theta^2)E_\theta\hat{\sigma}^2.$$

If B is sufficiently small, and/or if θ is such that σ_θ^2 is sufficiently large, it is true in many cases (as illustrated below) that $(1/\sigma_\theta^2)E_\theta\hat{\sigma}^2 \doteq 1$ and that $Bm/\sigma_\theta^2 \doteq 0$, and hence that $R(\theta) \doteq 1$; in such cases, for the indicated range of θ, no appreciable improvements in efficiency are possible even by resort to fully sequential estimators.

Such estimators have been found and investigated quantitatively for a number of common problems. These results are summarized in the following paragraphs.

4.1 Poisson Mean. If X has the Poisson distribution $f(x, \theta) = e^{-\theta}\theta^x / x!$ for $x = 0, 1, \cdots$, we may take

$$\hat{\sigma}^2 = \hat{\sigma}^2(x_1, \cdots x_m) = \left(\sum_1^m x_i + 1\right) \Big/ m,$$

since $y = \sum_1^m x_i$ has the Poisson distribution $f(y, m\theta) = e^{-m\theta}(m\theta)^y/y!$, $y = 0, 1, 2, \cdots$, and

$$E_\theta(1/\hat{\sigma}^2) = m \sum_{y=0}^\infty f(y, m\theta)/(y + 1) = me^{-m\theta} \sum_{y=0}^\infty (m\theta)^y/(y + 1)!$$

$$= (e^{-m\theta}/\theta) \sum_{y=0}^\infty (m\theta)^{y+1}/(y + 1)! = (1 - e^{-m\theta})/\theta < 1/\theta = 1/\sigma_\theta^2.$$

When the second sample size is determined by $n = \hat{\sigma}^2/B = (y + 1)/mB$, the expected total sample size is

$$E_\theta(N') = m + E_\theta(n) = m + (1/mB)E_\theta(y + 1) = m + (m\theta + 1)/mB.$$

This is minimized by taking $m \doteq 1/B^{1/2}$, regardless of the value of θ. (In other examples, an optimal first sample size is not so simple to determine.) Then $E_\theta(N') = \theta/B + 2/B^{1/2}$.

This estimator has efficiency bound $R(\theta)$ given by $1/R(\theta) = 1 + 2B^{1/2}/\theta$. If for example, $\theta = 8B^{1/2}$, then $R(\theta) = 0.8$, and a decrease of at most

$$100(1 - R(\theta))\% = 20\%$$

in $E_\theta(N')$ might be possible by resort to some (unknown) more refined sequential procedure; for $\theta \gg 8B^{1/2}$, the possible gains are negligible.

An alternative two-stage estimator (of the mean of a Poisson process) employing "inverse" sampling in the first stage, given in [10], has exactly variance B, but can be shown to be less efficient.

The following discussion illustrates that such estimators can have approximately normal distributions. For any fixed $\theta > 0$, $B > 0$, and k, we may write $\text{Prob}\{(T - \theta)/B^{1/2} < k\} = E_N U(N, \theta, k, B)$, where

$$U(N, \theta, k, B) = \text{Prob}\{(T - \theta)/B^{1/2} < k \mid N\}.$$

For sufficiently large fixed N,

$$U(N, \theta, k, B) = \Pr\{(T - \theta)/(\theta/N)^{1/2} < kB^{1/2}/(\theta/N)^{1/2}\} \doteq \Phi(k(BN/\theta)^{1/2}),$$

where $\Phi(u)$ is the standard normal c.d.f. As $B \to 0$, the random variable

$$\Phi(k(BN/\theta)^{1/2})$$

converges in probability to the constant $\Phi(k)$, as does the random variable

$U(N, \theta, k, B)$ with N random. Since $0 \leqq U \leqq 1$, we have

$$E_N(U) = \text{Prob}\{(T - \theta)/B^{1/2} < k\} \to \Phi(k)$$

as B decreases, proving the asymptotic normality of T.

4.2 Binomial Mean. If X has the binomial distribution

$$f(x, \theta) = \theta^x(1 - \theta)^{1-x}, \qquad x = 0, 1$$

we may take

$$\hat{\sigma}^2 = (1 - 2^{-(m+1)})\left(\sum_1^m x_i + 1\right)\left(m + 1 - \sum_1^m x_i\right)/(m + 1)(m + 2),$$

since (by a calculation similar to that in the preceding section)

$$E_\theta(1/\hat{\sigma}^2) = (1 - \theta^{m+2} - (1 - \theta)^{m+2})/(1 - 2^{-(m+1)})\theta(1 - \theta)$$
$$\leqq 1/\theta(1 - \theta) = 1/\sigma_\theta^2.$$

The expected total sample size is

$$E_\theta(N') = m + (1/B)E_\theta\hat{\sigma}^2 = m + (1 - 2^{-(m+1)})\,\frac{m(m-1)\theta(1-\theta) + m + 1}{B(m+1)(m+2)}$$
$$= m + (1 - 2^{-(m+1)})\,\frac{1}{B(m+2)} + \theta(1 - \theta)\left(1 - \frac{4m+2}{(m+1)(m+2)}\right).$$

The latter expression does not yield a minimizing value of m independent of the unknown θ, but for any chosen B and guessed value θ a minimizing value

$$m = m(\theta, B)$$

can be found by numerical solution of the equation $\frac{\partial}{\partial\theta} E_\theta(N') = 0$. Table 1 provides some such values.

TABLE 1

Best Binomial First Sample Sizes $m(\theta, B)$

θ	B			
	$(0.05)^2$	$(0.02)^2$	$(0.01)^2$	$(0.005)^2$
0.5	0	0	0	0
0.4 or 0.6	0	0	26	47
0.3 or 0.7	0	20	40	81
0.2 or 0.8	11	29	59	119
0.0 or 1.0	18	48	98	198

The value $m = 0$ indicates use of a single sample procedure with $n = 1/4B$ observations. However, calculations of $E_\theta(N')$ for various values of B and m, such as those given in Table 2, indicate that for $B \leqq (0.05)^2$ a choice of m such

as $m(0.2, B)$ provides appreciable savings as compared with $m = 0$ over a wide range of θ at the cost of a relatively small loss as compared with use of a best value $m(\theta', B)$ based on any guessed value θ' which happens to be correct.

TABLE 2

Values of $E_\theta(N')$ for Binomial Estimates

(a) $B = (0.05)^2$

θ	m		
	0	11	18
0.5	100	112.3	118.5
0.4 or 0.6	100	109.4	115.3
0.3 or 0.7	100	101.0	105.6
0.2 or 0.8	100	86.9	89.5
0.1 or 0.9	100	67.1	67.0
0.0 or 1.0	100	41.8	38.0

(b) $B = (0.005)^2$

θ	m				
	0	47	81	119	198
0.5	10,000	10,056	10,084	10,120	10,199
0.4 or 0.6	10,000	9,688	9,703	9,734	9,806
0.3 or 0.7	10,000	8,585	8,561	8,573	8,630
0.2 or 0.8	10,000	6,746	6,656	6,639	6,670
0.1 or 0.9	10,000	4,173	3,990	3,931	3,926
0.0 or 1.0	10,000	863	563	450	398

The efficiency bound $R(\theta)$ of such estimators is given by

$$1/R(\theta) = Bm/\theta(1 - \theta) + (1 - 2^{-(m+1)})$$
$$\cdot\{1/[(m + 2)\theta(1 - \theta)] + 1 - (4m + 2)/[(m + 1)(m + 2)]\};$$

For any given B, the values of m to be considered are $0 \leqq m \leqq m(0, B)$. If we take $m = m(0, B) = 1/B^{1/2} - 2 \doteq 1/B^{1/2}$ (for $B \leqq (0.05)^2$), we have

$$1/R(\theta) \doteq 1 + B^{1/2}(2/\theta(1 - \theta) - 4).$$

For any B, as $\theta \to 0$ or 1, $R(\theta) \to 0$; but these are values of θ for which the lower bound σ_θ^2/B on $E_\theta(N')$ cannot be attained by any estimator with the desired properties. For any fixed θ, $0 < \theta < 1$, as $B \to 0$, $R(\theta) \to 1$; thus the efficiency of $\hat{\theta}$ cannot be much improved upon when high precision is required. Analogous statements hold if we take, for example, $m(0.2, B)$.

The formula for the second sample size is

$$n = c\left(\sum_1^m x_i + 1\right)\left(m + 1 - \sum_1^m x_i\right),$$

where $c = c(B, m) = (1 - 2^{-(m+1)})/B(m + 1)(m + 2)$. Table 3 provides some values of $c(B, m)$.

TABLE 3

Values of $c(B, m)$ for Binomial Second Sample Sizes

m	B			
	$(0.05)^2$	$(0.02)^2$	$(0.01)^2$	$(0.005)^2$
10	3.03	18.94	75.76	303.04
15	1.47	9.19	36.76	147.04
20	0.87	5.41	21.64	86.56
25	0.57	3.56	14.24	56.96
30	0.40	2.52	10.08	40.32
35	0.30	1.88	7.52	30.08

The variance of $\hat{\theta}$ is

$$\mathrm{Var}_\theta(\hat{\theta}) = B(1 - \theta^{m+2} - (1 - \theta)^{m+2})/(1 - 2^{-(m+1)}) \leq B;$$

this is appreciably less than B only when θ is very near 0 or 1, provided m is not very small.

There exists, as in the Poisson case, a procedure employing "inverse" sampling to yield a binomial estimator having exactly constant variance as follows:

Let M be a fixed positive integer. Make successive independent Bernoulli trials (samples of size one) until min (total successes, total failures) $= M$. Let x be the number of trials up to and including the Mth success. Let y be the number of trials up to and including the Mth failure. Take an additional sample of size $n = M/B(x + y)$ and let z denote the number of additional successes observed. Then $\hat{\theta} = Bz(x + y)/M$ is an unbiased estimator of θ having variance exactly equal to B. It seems clear that the expected sample size will be larger for this "inverse" sampling plan, and that as a practical matter exactly prescribed variance would seldom be worth the cost in additional observations.

4.3 Hypergeometric mean. In a finite population of known size M, let $\theta = D/M$ be the unknown proportion of items having a given trait, e.g. being defective. Let X denote the number of defectives in a first sample of size m, $n = n(x)$ the size of a second sample, and Y the number of defectives in the second sample. (All sampling is without replacement.) Then it is readily verified that an unbiased estimator of θ is

$$\hat{\theta} = [(Y(M - m)/n) + X]/M.$$

This estimator will have variance bounded by B if we take

$$n = n(x) = \frac{(M - m)^2(x + 1)(m - x + 1)}{BM^2(m + 1)(m + 2) + (x + 1)(m - x + 1)(M - m)}$$

since

$$\mathrm{var}(\hat{\theta} \mid x) = \frac{(D - x)(M - m - D + x)(M - m - n)}{(M - m - 1)M^2 n}$$

and hence

$$
\begin{aligned}
\operatorname{var} \hat{\theta} &= E \operatorname{var}(\hat{\theta} \mid x) \\
&= \sum_{x=a}^{b} \frac{(D-x)(M-m-D+x)(M-m-n)}{(M-m-1)M^2 n} \binom{D}{x}\binom{M-D}{m-x} \Big/ \binom{M}{m} \\
&= \sum_{x=a}^{b} \frac{(M-m-n)(x+1)(m-x+1)(M-m)}{M^2 n(m+1)(m+2)} \\
&\qquad \cdot \binom{D}{x+1}\binom{M-D}{m-x+1} \Big/ \binom{M}{m+2} \\
&= \sum_{w=a+1}^{b+1} \frac{(M-m-n)w(m+2-w)(M-m)}{M^2 n(m+1)(m+2)} \\
&\qquad \cdot \binom{D}{w}\binom{M-D}{m+2-w} \Big/ \binom{M}{m+2} \\
&= B \sum_{w=a+1}^{b+1} \binom{D}{w}\binom{M-D}{m+2-w} \Big/ \binom{M}{m+2} \leqq B .
\end{aligned}
$$

Exact results on expected sample sizes are not available, but an indication of the possible savings is given by regarding the results in the binomial case to be limits approached as M becomes infinite. Additional information is given by the range of $n(x)$: for m even,

$$
\begin{aligned}
m + \frac{(M-m)^2}{M^2 B(m+2) + (M-m)} \leqq m + n(x) = N' \leqq m& \\
+ \frac{(M-m)^2(m+3)^2}{4BM^2(m+1)(m+2) + (M+3)^2(M-m)}&.
\end{aligned}
$$

With a single sample of size r, the best unbiased estimate has variance not exceeding B for all θ provided r is at least $M/(4B(M-1)+1)$. If for example $M = 1{,}000$, $B = .0001$, and $m = 80$, then $r = 714$ while $173 \leqq m + n(x) \leqq 728$; when $\theta = 0$ or 1, the two-stage estimate saves 541 observations (76%), while when $\theta = \frac{1}{2}$ its maximum sample size exceeds r by less than 14 observations (2%).

4.4 Mean of a Normal Distribution with Unknown Variance. If X has the normal density function $f(x, \theta, \sigma) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp (x - \theta)^2/2\sigma^2)$, with σ^2 unknown, we may apply the present method with an advantageous modification based on the independence of $\bar{x}_1 = (1/m)\sum_{i=1}^{m} x_i$ and

$$
s^2 = (1/(m-1)) \sum_{i=1}^{m} (x_i - \bar{x})^2.
$$

Take any $m > 3$, let $\hat{\sigma}^2 = (m-1)s^2/(m-3)$, $n = \max [\theta, \hat{\sigma}^2/B - m]$, and $\hat{\theta} = (1/(m+n)) \sum_{i=1}^{m+n} x_i$. It is easily verified that the latter estimator is unbiased and has variance not exceeding B. For most purposes Stein's procedure [1] which gives confidence intervals of prescribed length will probably be preferred; optimal choice of m for this procedure has been extensively investigated [7] and [17].

4.5 Estimation of a Scale Parameter. Let X have a density function

$$f(x, \theta) = (1/\theta)g(x/\theta), \qquad x \geqq 0,$$

and $f(x, \theta) = 0$ otherwise, where $g(u)$ is a known function with

$$c_1 = \int_0^\infty ug(u)\,du = E(X \mid \theta = 1),$$

$$c_2 = \int_0^\infty u^2 g(u)\,du = E(X^2 \mid \theta = 1) < \infty, \qquad \text{and}$$

$$c_3 = \int_0^\infty u^{-2} g(u)\,du = E\,(1/X^2 \mid \theta = 1) < \infty.$$

Then $E(X/c_1) = \theta$, $\text{Var}\,(X/c_1) = \theta^2(c_2/c_1^2 - 1) = \sigma^2$, say, and $E(1/X^2) = c_3/\theta^2$. Letting $\hat{\sigma}^2 = mc_3(c_2/c_1^2 - 1)[\sum_1^m 1/x_i^2]^{-1}$, we have

$$E(1/\hat{\sigma}^2) = 1/[c_3(c_2/c_1^2 - 1)]E(1/X^2) = 1/\sigma^2.$$

Thus an unbiased estimator of the scale-parameter θ, having variance B, is $\hat{\theta} = \sum_{i=m+1}^{m+n} x_i/c_1 n$. The choice of m may be made so as to minimize, at any guessed value of θ,

$$E_\theta(N') = m + E_\theta(n) = m + E_\theta(\hat{\sigma}^2)/B.$$

For any specific density function $f(x, \theta)$, it may be possible to find an estimate $\tilde{\sigma}^2$ preferable to the $\hat{\sigma}^2$ given above: $\tilde{\sigma}^2 = \tilde{\sigma}^2(x_1, \cdots x_m)$ is clearly preferable to $\hat{\sigma}^2$ if it (has the essential property $E(1/\tilde{\sigma}^2) \leqq 1/\sigma^2$ and also) makes $\hat{\theta}$ more efficient, that is, if $E(\tilde{\sigma}^2) < E(\hat{\sigma}^2)$. (This remark may be applied also to the estimators $\hat{\sigma}^2$ discussed in other sections of this paper.)

For example, if X has the gamma density

$$f(x, \theta) = (1/\theta\alpha!)(x/\theta)^\alpha e^{-x/\theta}, \qquad x \geqq 0,$$

where α is known, $\alpha > -1$, then

$$c_1 = \frac{1}{\alpha!}\int_0^\infty u^{\alpha+1} e^{-u}\,du = \alpha + 1,$$

$$c_2 = \frac{1}{\alpha!}\int_0^\infty u^{\alpha+2} e^{-u}\,du = (\alpha + 1)(\alpha + 2),$$

and, provided we require $\alpha > 1$,

$$c_3 = \frac{1}{\alpha!}\int_0^\infty u^{\alpha-2} e^{-u}\,du = 1/\alpha(\alpha - 1).$$

Thus $\hat{\sigma}^2 = m[(\alpha - 1)\alpha(\alpha + 1)\sum_{i=1}^m 1/x_i^2]^{-1}$. The evaluation of $E_\theta(\hat{\sigma}^2)$, required to compute $E_\theta(N')$, appears difficult for $m > 1$ and has not been carried out. For $m = 1$,

$$E_\theta(\hat{\sigma}^2) = E_\theta(X^2)/(\alpha - 1)\alpha(\alpha + 1) = \theta^2 c_2/(\alpha - 1)\alpha(\alpha + 1)$$
$$= \theta^2(\alpha + 2)/\alpha(\alpha - 1),$$

and $E_\theta(N') = 1 + \theta^2(\alpha + 2)/\alpha(\alpha - 1)B$. Presumably the estimator

$$\tilde{\sigma}^2 = K\left(\sum_1^m x_i\right)^2,$$

where K is such that $E_\theta(1/\tilde{\sigma}^2) = 1/\sigma^2$, is preferable to $\hat{\sigma}^2$ in this example, since $\tilde{\sigma}^2$ is a function of the sufficient statistic $\sum_1^m x_i$ (based on the first sample) and $\hat{\sigma}^2$ is not. Results of Ghurye [15] lend support to this conjecture. An estimator based on $\tilde{\sigma}^2$ is given in Section 5.3 below.

The estimation of any given power θ^p of a scale-parameter θ, $p = \pm 1, \pm 2, \cdots$, may be treated similarly.

5. Other estimation problems.

5.1 Variance of a normal distribution with unknown mean. Let X have the normal density function with unknown mean and variance as in Section 4.4, but let θ now denote the unknown *variance*. For any $m > 5$, let

$$n = 2s^2(m - 1)^2/B(m - 3)(m - 5) + 1,$$

where s^2 is the first sample variance defined as above. Then it is readily verified that an unbiased estimator of θ, with variance not exceeding B, is given by the second sample variance.

$$\hat{\theta} = \sum_{m+1}^{m+n}\left(x_i - \frac{1}{n}\sum_{m+1}^{m+n} x_j\right)^2 \bigg/ (n - 1),$$

and that

$$E_\theta(N') = m + 1 + \frac{2(m + 1)(m - 1)\theta^2}{(m - 3)(m - 5)B}.$$

For given B and a guessed value of θ, m may be chosen so as to minimize $E_\theta(N')$.

5.2 Estimation of a "between classes" variance component. Consider the usual assumptions for a one-way analysis of variance, with n observations from each of k classes: $Y_{ij} = \mu + c_i + e_{ij}$, $i = 1, \cdots k$, $j = 1, \cdots n$, with μ an unknown constant, and the c_i's and e_{ij}'s all independently normally distributed with means zero and unknown variances

$$\text{var}(c_i) = \sigma_0^2, \qquad \text{var}(e_{ij}) = \sigma^2, \qquad i = 1, \cdots k, j = 1, \cdots n.$$

The usual between classes mean square s_0^2 has expected value $\sigma^2 + n\sigma_0^2$ and $k - 1$ degrees of freedom. The usual within classes mean square s^2 has expected value σ^2 and $k(n - 1)$ degrees of freedom. Then $(s_0^2 - s^2)/n$ is an unbiased estimator of σ_0^2, with variance

$$2[(\sigma^2 + n\sigma_0^2)^2/(k - 1) + \sigma^4/k(n - 1)]/n^2$$

when k and n are fixed.

Alternatively, suppose a first sample of r classes and n observations per class has been taken. Let T_0^2 and T^2 respectively denote the between and within classes mean squares, based respectively on $\nu_0 = r - 1 > 4$ and $\nu = r(n - 1) > 4$

degrees of freedom. Then it is easily verified that

$$E(\nu_0 - 2)(\nu_0 - 4)/\nu_0^2 T_0^4 = 1/(\sigma^2 + n\sigma_0^2)^2$$

and

$$E(\nu - 2)(\nu - 4)/\nu^2 T^4 = 1/\sigma^4.$$

This leads to the choice of k defined by $k = \max(2, k', k'')$ where

$$k' = 1 + 2T_0^4\nu_0^2/(\nu_0 - 2)(\nu_0 - 4)(B - b)$$
$$k'' = 2T^4\nu^2/(\nu - 2)(\nu - 4)bn^2(n - 1)$$

and b is any constant, $0 < b < B$. To see that with k so defined, the sampling variance of $\hat{\sigma}_0^2$ is less than B, observe that

$$\begin{aligned} B &= 2[(\sigma^2 + n\sigma_0^2)^2 E(1/(k' - 1)) + (\sigma^4/(n - 1))E(1/k'')]/n^2 \\ &\geq 2[(\sigma^2 + n\sigma_0^2)^2 E(1/(k - 1)) + (\sigma^4/(n - 1))E(1/k)]/n^2 \\ &= \operatorname{var}(\hat{\sigma}_0^2). \end{aligned}$$

The choice of n would ordinarily be influenced by practical limitations on the experiment, and the choice of both n and b could also be governed by an a priori estimate of σ^2/σ_0^2.

An alternative approach to the present problem is to apply twice the method of the preceding Section 5.1 as follows: Estimate $(\sigma_0^2 + \sigma^2)$ by a two-stage estimator s_1^2 having variance not exceeding $B_1 < B$, based on observations $Y_{11}, Y_{21}, \cdots Y_{m,1}Y_{(m+1),1}, \cdots Y_{(m+n),1}$, so that only one observation is taken from each class. Secondly, estimate σ^2 by a two-stage estimator s_2^2 having variance not exceeding B_2, where $B_2 = B - B_1$, based on additional observations within any one class (or on additional "within degrees of freedom" from several classes). Then $s^2 = s_1^2 - s_2^2$ is the required estimate, for $E(s^2) = \sigma^2$, and

$$\operatorname{var}(s^2) \leq B_1 + B_2 = B.$$

Rules for optimal choice of B_1, and comparisons with the preceding method, remain to be developed.

5.3 Scale parameter of a gamma distribution. If X has the Gamma density defined in Section 4.5 above,

$$\operatorname{Var}(X/c_1) = \operatorname{Var}(X/(\alpha + 1)) = \theta^2(c_2/c_1^2 - 1) = \theta^2/(\alpha + 1) = \sigma^2,$$

and we may take

$$\hat{\sigma}^2 = \left(\sum_1^m x_i\right)^2/(\alpha + 1)(m\alpha + m - 2)(m\alpha + m - 1),$$

for all α and m such that $(m\alpha + m - 2) > 0$. This gives $E(1/\hat{\sigma}^2) = 1/\sigma^2$, and $E(\hat{\sigma}^2) = \theta^2(m\alpha + m + 1)(m\alpha + m)/(\alpha + 1)(m\alpha + m - 1)(m\alpha + m - 2)$. For any guessed value of θ, m may be chosen, subject to $m > 2/(\alpha + 1)$, so as to minimize $E_\theta(N') = m + E_\theta(\hat{\sigma}^2)/B$.

A modification analogous to that in Section 4.5 above, replacing $\sum x_1$ by

$(\sum x_i)^p$ throughout, with a corresponding modification of constants, gives an estimator of θ^p with variance B.

6. Applications to achieve homoscedasticity. Many standard techniques for comparing means, related to the analysis of variance (Model I), are seriously dependent for validity on the assumption that observations have (approximately) equal variances, but much less seriously dependent on the usual assumption of normality (see, for example, [8] and references therein). It is frequently desired to apply such methods to means of observations having some of the distributions considered in Section 4 above; but in such cases the unknown variances are functions of the unknown means of observations, and hence the assumption of equal variances generally holds only when the unknown means happen to be equal.

The methods of the present paper provide a way of meeting this difficulty which may be considered in cases where it is feasible to use a two-stage sampling method providing (approximately) a common prescribed variance B for the observation in each cell of any Model I experimental design. Techniques related to analysis of variance will be used taking, formally, the case of an infinite number of degrees of freedom for the error mean square; the latter, of course, will not be calculated from data, but the known variance B will be used instead. The methods which are usually considered for meeting this difficulty are variance-stabilizing transformations of the observations (see, for example, [9]). Concerning the relative advantages and disadvantages of these approaches, it should be noted that the goal of (a) variance-stabilization for application of standard inference techniques is usually of interest simultaneously with certain goals of (b) precision of estimation (or power of tests), (c) efficient utilization of data obtained, and (d) simplicity of interpretation. Concerning (d), use of the methods of this paper offers some advantages over use of transformations since the former provides inferences directly about the means of interest with prescribed precision on their original scale, rather than inferences about functions of those means (e.g., $E(\sin^{-1}(x/n)^{\frac{1}{2}})$ in the binomial case) which are often harder to interpret and perhaps less meaningful. Furthermore, the latter estimators lose their constant-precision property when interpreted in the original units of the parameters.

In cases like the Poisson there is no single-sample procedure which provides even bounded, let alone prescribed, precision in the original scale. Hence if such prescribed precision is one goal of interest, sequential methods more or less like those of this paper are required, and the simultaneous achievement of simplicity, and of exactly or approximately known common variances of estimators of means, may be regarded as convenient desirable by-products of the method.

In cases like the binomial, the goals of bounded precision and homoscedasticity are attainable by use of transformed single-sample estimates. In the binomial case, we have seen above that when high constant precision is desired, the two-sample estimate is on the whole rather efficient, and in this case again affords the properties of homoscedasticity and simplicity. If only low precision is required, there is some conflict between the goals mentioned. For example, for

binomial estimation with $B = (.05)^2$ it was shown above that a first-sample size of $m \geqq 20$ gives an inefficient estimate, but $m \geqq 20$ is required for a good degree of homoscedasticity. In such cases efficiency considerations may be weighed against considerations of simplicity of application and interpretation. If it can be assumed that $.2 \leqq \theta \leqq .8$, then $m \geqq 10$ suffices to give a variation of at most 7% in variances of $\hat{\theta}$'s. If it can be assumed that $.1 \leqq \theta \leqq .9$, the variation is at most 10% if $m \geqq 20$.

7. Acknowledgements. The authors are grateful to Dr. Donald Guthrie and to the Applied Mathematics and Statistics Laboratory of Stanford University for permission to reprint some of the tables above from [11] and [12]. Mr. B. B. Bhattacharya [14] of the Indian Statistical Institute independently found some estimates of the type given in this paper.

REFERENCES

[1] Stein, C. M., "A two-sample test for a linear hypothesis whose power is independent of the variance," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 243–258.

[2] Cox, D. R., "A note on the sequential estimation of means," *Proc. Camb. Phil. Soc.* Vol. 48 (1952), pp. 447–450.

[3] Cox, D. R., "Estimation by double sampling," *Biometrika*, Vol. 39 (1952), pp. 217–227.

[4] Anscombe, F. J., "Sequential estimation," *J. Roy. Stat. Soc., Ser. B*, Vol. 15 (1953), pp. 1–29.

[5] Wolfowitz, J., "The efficiency of sequential estimates and Wald's equation for sequential processes," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 215–230.

[6] Birnbaum, A., "Sequential probability ratio confidence sets" (Abstract), *Ann. Math. Stat.*, Vol. 24 (1953), p. 686.

[7] Seelbinder, B. M., "On Stein's two-stage sampling scheme," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 640–647.

[8] Box, G. E. P., "Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 290–302; and "—II. Effects of inequality of variance and of correlation between errors in the two-way classification," *ibid.*, pp. 484–498.

[9] Freeman, M. F., and Tukey, J. W., "Transformations related to the angular and the square root," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 607–611.

[10] Birnbaum, A., "Statistical methods for Poisson processes and exponential populations," *J. Amer. Stat. Assn.*, Vol. 49 (1954) pp. 254–266.

[11] Guthrie, D., "Some two-sample procedures," unpublished M. A. Thesis, Columbia University, 1955.

[12] Birnbaum, A., and Guthrie, D., "Tables for estimating a proportion or a Poisson mean with prescribed precision," *Technical Report No. 25*, Jan. 18, 1956, Applied Mathematics and Statistics Laboratory, Stanford University.

[13] Healy, W. C., "Multiple sampling to estimate a proportion," *O.O.R. Technical Report*, Dec. 15, 1954, Statistical Research Laboratory, University of Illinois.

[14] Bhattacharya, B. B. Unpublished note on sequential estimation.

[15] Ghurye, S. G., "Note on sufficient statistics and two-stage procedures," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 155–166.

[16] Graybill, Franklin A., "Determining sample size for a specified width confidence interval," Ann. Math. Stat., Vol. 29 (1958), pp. 282–287.

[17] Moshman, Jack, "A method for selecting the size of the initial sample in Stein's two sample procedure," *Ann. Math. Stat.*, Vol. 29 (1958) pp. 1271–1275.

A MONOTONICITY PROPERTY OF THE SEQUENTIAL PROBABILITY RATIO TEST[1]

By Robert A. Wijsman

University of Illinois

0. Summary. Using the basic inequalities (1) it is shown that, if, in a sequential probability ratio test, the upper stopping bound is increased and the lower stopping bound decreased, and if the new test is not equivalent to the old one, then at least one of the error probabilities is decreased. This implies the monotonicity result of Weiss [5] in the continuous case, and the uniqueness result of Anderson and Friedman [1] in the general case. The relation of the monotonicity property to the optimum property and the uniqueness of sequential probability ratio tests is discussed.

The monotonicity property is a consequence of the following stronger result. Let the old and new tests be given by the stopping bounds (B', A') and (B, A), respectively, with $B < B' < A' < A$; let (α_1', α_2') and (α_1, α_2) be the error probabilities and $\Delta\alpha_i = \alpha_i - \alpha_i'$ the changes in the error probabilities; then the vector $(\Delta\alpha_1, \Delta\alpha_2)$ is restricted to a cone consisting of the 3rd quadrant, plus the part of the 2nd quadrant where $-\Delta\alpha_2/\Delta\alpha_1 < B$, plus the part of the 4th quadrant where $-\Delta\alpha_2/\Delta\alpha_1 > A$. Another consequence of this result is that (α_1, α_2) cannot lie in the closed triangle with vertices (α_1', α_2'), $(0, 1)$ and $(1, 0)$. Finally, the following monotonicity property follows: If the lower stopping bound is fixed and the upper stopping bound increased, then $\alpha_1/(1 - \alpha_2)$ decreases monotonically. The same holds for $\alpha_2/(1 - \alpha_1)$ if the upper stopping bound is held fixed and the lower stopping bound decreased.

1. Introduction and discussion. We consider Wald's sequential probability ratio test [3] with upper stopping bound A and lower stopping bound B. It is usually assumed that $B < 1 < A$, but no such restriction will be made in this paper. Weiss [5] has shown, under certain continuity assumptions, that, if A and B are separated in such a way that one of the error probabilities remains constant, then the other error probability decreases monotonically. This is a very useful result, since it not only provides a uniqueness proof, but also it shows that there exists a test of given strength if and only if the error probability vector lies in a certain set [6]. In this paper a monotonicity property will be proved which makes no assumptions as regards to the probability distributions (other than that they be non-degenerate) and which include Weiss' result as a special case. The monotonicity property, stated and proved in Section 2, can be described as follows: if the upper stopping bound of a sequential probability ratio test is increased and the lower stopping bound decreased, then at least one of the error probabilities decreases, unless the new test is equivalent to the old one, in

Received September 26, 1959; revised January 25, 1960.

[1] Work supported by the National Science Foundation, grant NSF G-9104.

which case the error probabilities are, of course, unchanged. (Two tests will be called *equivalent*, more or less following [1], if their sample sequences differ on a set of probability 0 under both distributions.) Weiss' result is obtained as a particular case by specifying the distributions to be continuous, with positive probabilities in non-degenerate intervals, and by reading the conclusion: then if one of the error probabilities is fixed, the other decreases.

Before proving the indicated monotonicity property, its relation to the uniqueness and to the optimum property [4] of sequential probability ratio tests will be discussed. In [1] it is shown how the optimum property can be used to prove uniqueness, i.e. the fact that two sequential probability ratio tests with the same error probabilities are equivalent. The restriction $B < 1 < A$ had to be made, though, since the optimum property had been proved only under this condition. Actually, this restriction is unnecessary. It will be indicated in a future paper [2] that any sequential probability ratio test has the optimum property among all tests which take at least one observation. In particular, then, every sequential probability ratio test has the optimum property among all sequential probability ratio tests, which is all that is needed in the uniqueness proof in [1]. This kind of optimum property will be labeled *restricted* in the following.

First of all it will be shown now that the restricted optimum property and the monotonicity property are equivalent. The following notation and terminology will be used: the error probabilities corresponding to the two distributions under consideration are denoted by α_i, $i = 1, 2$; the expected sample sizes are ν_i; in passing from one test to another, $\Delta\alpha_i$ and $\Delta\nu_i$ denote the changes in the α_i and ν_i; a test will be called *inadmissible* if there exists another test such that $\Delta\alpha_i \leqq 0$, $\Delta\nu_i \leqq 0$, $i = 1, 2$, with strict inequality in at least one of the four. Obviously, the optimum property implies admissibility, and the restricted optimum property implies restricted admissibility, i.e. admissibility within the class of sequential probability ratio tests. Consider a sequential probability ratio test (B, A) and another, (B^*, A^*), with $B^* \leqq B < A \leqq A^*$. Unless the two tests are equivalent, we have $\Delta\nu_i > 0$ for both i (see Section 2 for support of this statement, and similar ones to follow). Assume the restricted optimum property. This implies restricted admissibility, and this implies that $\Delta\alpha_i < 0$ for at least one i. In other words, one of the α_i has to decrease, which is the monotonicity property. Conversely, assume the monotonicity property, and compare tests (B, A) and (B^*, A^*), which are supposed to be not equivalent and for which $\Delta\alpha_i \leqq 0$ for both i. Then we cannot have $B^* \leqq B$, $A^* \leqq A$, for in that case $\Delta\alpha_1 > 0$ and $\Delta\alpha_2 < 0$. Similarly, $B^* \geqq B$, $A^* \geqq A$ is excluded. Also $B < B^* < A^* < A$ is excluded since otherwise, by the monotonicity property, one of the $\Delta\alpha_i$ would be positive. Hence the only remaining possibility is $B^* < B < A < A^*$, which implies $\Delta\nu_i > 0$ for both i, i.e. the optimum property.

Secondly, the monotonicity property implies uniqueness. For, if the stopping bounds are changed in the same direction, then both error probabilities change (in opposite directions), whereas, if the stopping bounds are changed in opposite directions, then according to the monotonicity property at least one of the error probabilities changes; unless, of course, the two tests are equivalent.

It is true that a separate proof of the monotonicity property is not strictly necessary, since this property is a consequence of the optimum property. However, the optimum property is a rather deep theorem, requiring a sizable machinery for its proof, whereas the monotonicity property follows in an elementary way from the basic inequalities (1). Since two interesting properties of sequential probability ratio tests within their own class—the uniqueness and the restricted optimum property—are immediate consequences of the monotonicity property, it seems worth-while to prove the latter independently. Moreover, the methods used yield a stronger result, which does not follow from the optimum property and which has, besides the monotonicity property, some other interesting consequences. These further results are obtained in Section 3.

2. Statement and proof of the monotonicity property. Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables (or vectors) with common density p_i with respect to some sigma-finite measure. Here and in the following, i runs over 1 and 2, corresponding to the two hypotheses under consideration. The trivial case $p_1 = p_2$ a.e. will be excluded. Let Y_n be the probability ratio at the nth observation, i.e. $Y_n = \prod_{j=1}^{n} p_2(X_j)/p_1(X_j)$. If some stopping rule is defined, let N be the random number of observations. Of fundamental importance in what follows is the basic double inequality

$$aP_1(a < Y_N < b) \leqq P_2(a < Y_N < b) \leqq bP_1(a < Y_N < b) \tag{1}$$

for any real numbers a and b, including ∞. The strict inequality signs within the parentheses in (1) may be replaced by less-or-equal signs, and we will do so whenever this is convenient. For instance, the following inequalities will be considered special cases of (1):

$$P_2(Y_N \geqq a) \geqq aP_1(Y_N \geqq a) \tag{2}$$

$$P_2(Y_N \leqq b) \leqq bP_1(Y_N \leqq b). \tag{3}$$

These basic inequalities have been used already by Wald ([3], Section 3.2) and are briefly discussed there. Also Weiss [5] makes use of (3). An important consequence of (1) is that either

$$P_1(a < Y_N < b) = P_2(a < Y_N < b) = 0 \tag{4}$$

or

$$P_1(a < Y_N < b) > 0 \quad \text{and} \quad P_2(a < Y_N < b) > 0. \tag{5}$$

As an application, compare the sequential probability ratio tests (B, A) and (B, A^*), with $B < A < A^*$. In (4) and (5) identify a with A, b with A^*, and N with the random number of observations if test (B, A) is used. If (4) prevails, the two tests are clearly equivalent. If (5) prevails we can conclude $\Delta\alpha_1 < 0$, $\Delta\alpha_2 > 0$, $\Delta\nu_i > 0$ for both i. Similar conclusions can be drawn if both stopping bounds are changed, and these facts have already been used in the discussion in Section 1.

In a sequential probability ratio test with stopping bounds s and $t (s < t)$[2] and random number of observations N, the error probabilities α_i are functions of s and t,

$$(6) \qquad \alpha_1(s, t) = P_1(Y_N \geqq t) = 1 - P_1(Y_N \leqq s),$$

$$(7) \qquad \alpha_2(s, t) = P_2(Y_N \leqq s) = 1 - P_2(Y_N \geqq t).$$

It is convenient to introduce the functions U_i and V_i defined by

$$(8) \qquad U_i(s, t) = P_i(Y_N \leqq s)$$

$$(9) \qquad V_i(s, t) = P_i(Y_N \geqq t)$$

if s and t are the stopping bounds. We have

$$(10) \qquad U_i(s, t) + V_i(s, t) = 1,$$

and the relation between the α_i, U_i and V_i is simply

$$(11) \qquad \alpha_1(s, t) = V_1(s, t) \qquad \alpha_2(s, t) = U_2(s, t).$$

THEOREM 1. *Let (u, v) and (u', v') define two non-equivalent sequential probability ratio tests, with $0 < u \leqq u' < v' \leqq v < \infty$, and let $\Delta\alpha_i = \alpha_i(u, v) - \alpha_i(u', v')$, $i = 1, 2$. Then at least one of the $\Delta\alpha_i$ must be <0.*

PROOF. Let N be the random number of observations in the test (u', v'), and define $F_i(y) = P_i(Y_N \leqq y)$. Then, using (11), we have

$$(12) \qquad \Delta\alpha_1 = V_1(u, v) - V_1(u', v')$$

$$(13) \qquad \Delta\alpha_2 = U_2(u, v) - U_2(u', v').$$

We compute[3]

$$(14) \qquad V_1(u', v') = \int_v^\infty dF_1(y) + \int_{v'}^v dF_1(y)$$

$$(15) \qquad V_1(u, v) = \int_v^\infty dF_1(y) + \int_{v'}^v V_1\left(\frac{u}{y}, \frac{v}{y}\right) dF_1(y) + \int_u^{u'} V_1\left(\frac{u}{y}, \frac{v}{y}\right) dF_1(y).$$

In (15) we used the fact that the Y_{n+1}/Y_n are independent and identically distributed. Substitution into (12) and using (10) gives

$$(16) \qquad \Delta\alpha_1 = \int_u^{u'} V_1\left(\frac{u}{y}, \frac{v}{y}\right) dF_1(y) - \int_{v'}^v U_1\left(\frac{u}{y}, \frac{v}{y}\right) dF_1(y).$$

[2] For notational convenience we shall henceforth use lower case symbols instead of A and B for the stopping bounds.

[3] In (14) the lower limits on the integrals should, and the upper limits should not be included in the integrations. On the other hand, in (15) in the third integral on the right the lower limit u should not be included and the upper limit u' should. These facts have not been made explicit in the formulas, since they are inessential for the proof.

Similarly,

$$\Delta\alpha_2 = \int_{v'}^{v} U_2\left(\frac{u}{y}, \frac{v}{y}\right) dF_2(y) - \int_{u}^{u'} V_2\left(\frac{u}{y}, \frac{v}{y}\right) dF_2(y). \tag{17}$$

Suppose temporarily that $v' < v$. Then for y in the interval $[v', v)$ we have, using (8) and (3),

$$U_2\left(\frac{u}{y}, \frac{v}{y}\right) \leqq \frac{u}{y} U_1\left(\frac{u}{y}, \frac{v}{y}\right) \leqq \frac{u}{v'} U_1\left(\frac{u}{y}, \frac{v}{y}\right), \tag{18}$$

and, using (1),

$$dF_2(y) \leqq v\, dF_1(y) \tag{19}$$

so that

$$\int_{v'}^{v} U_2\left(\frac{u}{y}, \frac{v}{y}\right) dF_2(y) \leqq \frac{uv}{v'} \int_{v'}^{v} U_1\left(\frac{u}{y}, \frac{v}{y}\right) dF_1(y). \tag{20}$$

In (19), and therefore in (20), there is strict inequality unless

$$\int_{v'}^{v} dF_i(y) = 0 \qquad \text{for both } i. \tag{21}$$

If $v' = v$, (20) remains trivially true. Similarly,

$$\int_{u}^{u'} V_2\left(\frac{u}{y}, \frac{v}{y}\right) dF_2(y) \geqq \frac{uv}{u'} \int_{u}^{u'} V_1\left(\frac{u}{y}, \frac{v}{y}\right) dF_1(y), \tag{22}$$

and again this inequality is strict unless

$$\int_{u}^{u'} dF_i(y) = 0 \qquad \text{for both } i. \tag{23}$$

The tests are equivalent if and only if both (21) and (23) hold. Therefore, if the tests are not equivalent, then at least one of the inequalities in (20) and (22) is strict. Using (16), (17), (20) and (22), it is now easy to verify the following two inequalities:

$$uv\Delta\alpha_1 + v'\Delta\alpha_2 < 0 \tag{24}$$

$$uv\Delta\alpha_1 + u'\Delta\alpha_2 < 0. \tag{25}$$

The conclusion of Theorem 1 is, of course, an immediate consequence of either of the inequalities (24) and (25).

3. Strengthening of the result. Let $\Delta\alpha$ be the 2-vector whose components $\Delta\alpha_i$ are defined in Theorem 1. If the two tests are equivalent, then, of course, $\Delta\alpha = 0$. Otherwise, the conclusion of Theorem 1 states that $\Delta\alpha$ cannot lie in the set defined by $\Delta\alpha_i \geqq 0$ for both i, i.e. the (closed) 1st quadrant. In other words, $\Delta\alpha$ has to lie in the 2nd, 3rd or 4th quadrant. However, the inequalities (24) and (25) already claim something more: $\Delta\alpha$ is not only excluded from the 1st

quadrant but also from the part of the 2nd quadrant where $-\Delta\alpha_2/\Delta\alpha_1 \geqq uv/v'$ (using (24)) and from the part of the 4th quadrant where $-\Delta\alpha_2/\Delta\alpha_1 \leqq uv/u'$ (using (25)). What remains is a cone of angle $<\pi$. We shall show now that we can sharpen the bounds uv/v' and uv/u' for $-\Delta\alpha_2/\Delta\alpha_1$ to u and v, respectively. This will be the content of

THEOREM 2. *Under the same conditions as in Theorem* 1 *we have*

$$u\Delta\alpha_1 + \Delta\alpha_2 < 0 \tag{26}$$

$$v\Delta\alpha_1 + \Delta\alpha_2 < 0. \tag{27}$$

Before proving Theorem 2 we will indicate some of its consequences. Consider u', v' fixed and u, v varying, subject to $u < u' < v' < v$. Consider all possible $\Delta\alpha$. The cone given by (26) and (27) to which $\Delta\alpha$ is restricted depends on u and v. To obtain a fixed cone we remark that $-\Delta\alpha_2/\Delta\alpha_1 < u$ implies $-\Delta\alpha_2/\Delta\alpha_1 < u'$ and $-\Delta\alpha_2/\Delta\alpha_1 > v$ implies $-\Delta\alpha_2/\Delta\alpha_1 > v'$. Therefore, (26) and (27) imply

$$u'\Delta\alpha_1 + \Delta\alpha_2 < 0 \tag{28}$$

$$v'\Delta\alpha_1 + \Delta\alpha_2 < 0. \tag{29}$$

The inequalities (28) and (29) are less sharp than (26) and (27), but they do represent a fixed cone within which $\Delta\alpha$ is restricted as u and v vary. In fact, this cone is the union of all cones given by (26) and (27) as u and v vary.

We can also consider the $\alpha_1 - \alpha_2$ plane and see what happens to the vector of error probabilities as u', v' is fixed and u, v vary. The only portion of the plane which needs to be considered is the triangle $\alpha_i \geqq 0$, $\alpha_1 + \alpha_2 \leqq 1$. Let $\alpha_i = \alpha_i(u, v)$ $\alpha_i' = \alpha_i(u', v')$, and let $\alpha = (\alpha_1, \alpha_2)$, $\alpha' = (\alpha_1', \alpha_2')$. The inequalities (28) and (29) say that α lies in a cone with vertex α', containing the point (0, 0), and bounded by two lines with slopes $-u'$ and $-v'$. This cone does not contain any point of the triangle with vertices α', (0, 1) and (1, 0). To see this we only have to look at the slopes of the lines connecting α' with (0, 1) and (1, 0). The first is $-(1 - \alpha_2')/\alpha_1'$, the second $-\alpha_2'/(1 - \alpha_1')$. Now, using (2), we have $(1 - \alpha_2')/\alpha_1' \geqq v' > u'$, and, using (3), $\alpha_2'/(1 - \alpha_1') \leqq u' < v'$, which establishes the fact mentioned. Thus, α cannot lie in the closed triangle with vertices α', (0, 1) and (1, 0).

There is another consequence which is of enough interest in itself to state separately. We introduce the quantities

$$\beta_1 = \alpha_1/(1 - \alpha_2) \tag{30}$$

$$\beta_2 = \alpha_2/(1 - \alpha_1) \tag{31}$$

The quantities β_i' are defined in the same manner in terms of the α_i', and $\Delta\beta_i = \beta_i - \beta_i'$. Then β_1 is the tangent of the angle that the line through α and (0, 1) makes with the α_2-axis; β_2 has a similar interpretation. The result of the preceding paragraph, namely that α is excluded from the closed triangle with vertices α', (0, 1) and (1, 0), is then seen to be equivalent to

$$\Delta\alpha_1 < 0 \Rightarrow \Delta\beta_1 < 0 \tag{32}$$

$$(33) \qquad \Delta\alpha_2 < 0 \Rightarrow \Delta\beta_2 < 0.$$

This result can also be stated as

COROLLARY 1. *Under the same conditions as in Theorem 1, at least one of the $\Delta\beta_i$ must be < 0.*

Now $\Delta\alpha_1 < 0$ is in particular satisfied if $u = u'$, and $\Delta\alpha_2 < 0$ if $v = v'$. Using this, and referring to (32) and (33), we have

COROLLARY 2. *Let the β_i be defined by (30) and (31). If the lower stopping bound u of a sequential probability ratio test is fixed, β_1 is a monotonic non-increasing function of the upper stopping bound v. The function is strictly monotonic except in any point v for which there is a $v^* > v$ such that the tests (u, v) and (u, v^*) are equivalent. A completely analogous statement for β_2 is obtained by fixing v and decreasing u.*

Finally, we remark that Theorem 2 can be generalized slightly. So far we have considered only sequential probability ratio tests whose continuation region is an open interval. We can also consider a sequential probability ratio test whose continuation interval contains one or both of its endpoints. In Theorems 1 and 2 we shall then consider tests with continuation intervals I and I', where I has endpoints u, v, and I' has u', v', and such that $I' \subset I$. With this generalization the conclusion (26) and (27) remains valid, except that one of the inequalities may be an equality.

We proceed now with the proof of Theorem 2, which starts with (24) and (25). Notice that for very small changes from u' to u and v' to v we have almost $u/u' = 1$ and $v/v' = 1$ so that then (26) and (27) follow approximately from (24) and (25), respectively. The idea of the proof is to link the tests (u', v') and (u, v) by a chain of intermediate tests, each of which is close to the next one.

PROOF OF THEOREM 2. Consider the chain of tests (u', v'), (u, v') (u, v_1), $\cdots$, (u, v_n) in which $v_n = v$ and $v_1, \cdots, v_{n-1}$ is a sequence to be specified later. Put $(\Delta\alpha_i)_0 = \alpha_i(u, v') - \alpha_i(u', v')$ and $(\Delta\alpha_i)_k = \alpha_i(u, v_k) - \alpha_i(u, v_{k-1})$, $k = 1, \cdots, n$, where we identify v_0 with v'. In passing from (u', v') to (u, v') we have $(\Delta\alpha_1)_0 \geqq 0$, with strict inequality unless (u', v') and (u, v') are equivalent. Consequently, using (25) in the second inequality, $u(\Delta\alpha_1)_0 + (\Delta\alpha_2)_0 \leqq u(v/u')(\Delta\alpha_1)_0 + (\Delta\alpha_2)_0 \leqq 0$, with equality if and only if (u', v') and (u, v') are equivalent. Then there exists $\epsilon_0 > 0$ such that for all ϵ with $0 < \epsilon < \epsilon_0$ we have

$$(34) \qquad u(\Delta\alpha_1)_0 + (1 - \epsilon)(\Delta\alpha_2)_0 \leqq 0$$

with the same remark about equality as before. For fixed ϵ, $0 < \epsilon < \epsilon_0$, choose $v_1, \cdots, v_{n-1}$ in such a way that $1 - \epsilon < v_{k-1}/v_k < 1$, $k = 1, \cdots, n$. In passing from (u, v_{k-1}) to (u, v_k) we have $(\Delta\alpha_2)_k \geqq 0$ so that $u(\Delta\alpha_1)_k + (1 - \epsilon)(\Delta\alpha_2)_k \leqq u(\Delta\alpha_1)_k + (v_{k-1}/v_k)(\Delta\alpha_2)_k$. The right hand side of the last inequality is $\leqq 0$, by (24), with equality if and only if (u, v_{k-1}) and (u, v_k) are equivalent. We have established now

$$(35) \qquad u(\Delta\alpha_1)_k + (1 - \epsilon)(\Delta\alpha_2)_k \leqq 0, \qquad k = 0, 1, \cdots, n$$

(for $k = 0$ this was established as (34)). In (35) there is strict inequality for

at least one k, otherwise (u', v') and (u, v) would be equivalent. Adding the inequalities (35) for $k = 0, 1, \cdots, n$ yields

$$u\Delta\alpha_1 + (1 - \epsilon)\Delta\alpha_2 < 0. \tag{36}$$

Letting $\epsilon \to 0$ then gives the desired result (26). Inequality (27) is proved analogously, using a chain of tests (u', v'), (u', v), (u_1, v), $\cdots$, (u_n, v), with $u_n = u$.

REFERENCES

[1] T. W. Anderson and Milton Friedman, "A limitation of the optimum property of the sequential probability ratio test," No. 6 in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, Stanford University Press, Stanford, 1960,

[2] Donald L. Burkholder and Robert A. Wijsman, "Optimum property and inadmissibility of sequential tests," (in preparation).

[3] A. Wald, *Sequential Analysis*, New York, John Wiley and Sons, 1947.

[4] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 326–339.

[5] Lionel Weiss, "On the uniqueness of Wald sequential tests," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1178–1181.

[6] Robert A. Wijsman, "On the existence of Wald's sequential test," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 938–939, (abstract).

LOCALLY MOST POWERFUL RANK TESTS FOR TWO-SAMPLE PROBLEMS[1]

By Hirofumi Uzawa

Stanford University

1. Summary and Introduction. In order to solve nonparametric statistical problems, it is often found useful to apply those criteria of optimality which are employed in parametric problems. In the present paper, we are concerned with nonparametric two-sample problems of testing the null hypothesis that two populations have the same distribution against certain nonparametric alternative hypotheses, and generalize the parametric optimality conditions of locally most powerfulness. A rank test for a two-sample problem in this paper is called locally most powerful if it is locally most powerful against a one-parameter family of alternatives. A criterion is constructed by which it is possible to solve the problem whether or not a given rank test is locally most powerful for a two-sample problem in which the set of all possible pairs of cumulative distribution functions is convex and closed (in the weak* topology).

Let $X_1, \cdots, X_{n_1}$ be sample elements from the first population, $X_{n_1+1}, \cdots, X_n$ $(n = n_1 + n_2)$ from the second population, and the statistic[2] $Z_j = 0$ or 1, according to whether the jth smallest observation is from the first population or the second, $j = 1, \cdots, n$.

It will be shown that any locally most powerful rank test has the following form:

$$(*)\qquad \begin{cases} \text{Reject the null hypothesis if } \sum_j a_j Z_j > c \\ \text{Accept the null hypothesis if } \sum_j a_j Z_j < c \end{cases}$$

where $a_1, \cdots, a_n$ are constant numbers. For the two-sided two-sample problem, any rank test of the form (*) is locally most powerful. For the one-sided two-sample problem, a non-trivial rank test of the form (*) is locally most powerful if, and only if,

$$c_j = \left[1 \Big/ \binom{n-2}{j}\right] \sum_{s=j}^{n-2} \binom{s+1}{j+1} (a_{s+2} - \bar{a}), \qquad j = 0, 1, \cdots, n-2,$$

where

$$\bar{a} = \frac{1}{n}(a_1 + \cdots + a_n),$$

Received September 29, 1958; revised February 15, 1960.

[1] The present work was begun at the Summer Statistical Institute on Nonparametric Methods held at the University of Minnesota in June-August, 1958, under the sponsorship of the National Science Foundation and was completed with the support of an Office of Naval Research contract at Stanford University.

[2] The whole argument of the present paper is very much simplified by the use of the Z-statistics which are defined by I. R. Savage [6].

have all non-negative Hankel determinants (the precise definition of the Hankel determinants is given in Section 8 below).

For the symmetric two-sided two-sample problem, a non-trivial rank test of the form (*) is locally most powerful if, and only if,

$$\sum_{s=j+1}^{n} \binom{s-1}{j} (a_{n+1-s} - a_s) = 0, \qquad \text{for } j = 1, \cdots, n.$$

Finally, it will be shown that for the two-sample problem, in which the alternative hypothesis is that the expectation of the first cumulative distribution function with respect to the second distribution is not less than $\frac{1}{2}$, a non-trivial rank test of the form (*) is locally most powerful if, and only if,

$$\sum_{s=1}^{[(n+1)/2]} \left(\frac{n+1}{2} - s\right)(a_{n-s} - a_s) \geqq 0.$$

2. Two-Sample Problems. Suppose that there are two statistical populations with cumulative distribution functions $F(x)$ and $G(x)$, $-\infty < x < +\infty$. A two-sample problem is concerned with testing a null hypothesis H_0 against a certain alternative hypothesis H_1 based upon the observation of finite random samples $X_1, \cdots, X_{n_1}$ and $X_{n_1+1}, \cdots, X_n$ taken from populations F and G, respectively. We will confine ourselves here to the cases in which the sizes n_1 and $n_2 = n - n_1$ of random samples are fixed.

In the present paper, our main interest will be in the following two-sample problems:

Problem (I): Two-sided two-sample problem. Test the null hypothesis H_0 that two populations F and G have the same distribution: $F = G$, against the alternatives H_1 that two populations have different distributions: $F \neq G$.

Problem (II): One-sided two-sample problem. Test the same null hypothesis H_0 against the alternatives H_2 that the first population F is statistically smaller than the second population G:

$$F \geq G.$$

Problem (III): Symmetric two-sided two-sample problem. Test the null hypothesis H_0 that F and G are symmetric and identical against the alternatives H_3 that the two populations are both symmetric with the same median but are different.

Problem (IV): Test the null hypothesis H_0 against the alternatives H_4 that two populations have different distributions and the mean of the first F with respect to G is not greater than $\frac{1}{2}$:

$$F \neq G \quad \text{and} \quad \int F\, dG \geqq \frac{1}{2}.$$

For the sake of simplicity, we use the following notation: For two functions F and G,

$$F = G \quad \text{if} \quad F(x) = G(x) \quad \text{for all} \quad x,$$

$$F \geqq G \quad \text{if} \quad F(x) \geqq G(x) \quad \text{for all} \quad x,$$
$$F \geq G \quad \text{if} \quad F \geqq G \quad \text{but} \quad F \neq G.$$

It will always be assumed that $F(x)$ is strictly increasing and continuous on $-\infty < x < +\infty$.

3. Rank Tests. Let $X_1, \cdots, X_{n_1}$, and $X_{n_1+1}, \cdots, X_n$ be two random samples taken from populations F and G, respectively. We will confine our attention to rank tests which may be defined conveniently in terms of the following Z statistics

$$Z_j = \begin{cases} 0, & \text{if the } i\text{th smallest observation among } X_1, \cdots, X_n \text{ comes from population } F \\ 1, & \text{otherwise,} \end{cases}$$

$$j = 1, \cdots, n;$$

then any non-randomized rank test ϕ may be expressed by

$$\phi(X_1, \cdots, X_n) = \begin{cases} 0, & \text{if} \quad T(Z_1, \cdots, Z_n) < c \\ 1, & \text{if} \quad T(Z_1, \cdots, Z_n) > c, \end{cases} \tag{1}$$

where $T(z_1, \cdots, z_n)$ is a function defined on $z = (z_1, \cdots, z_n)$ with $z_j = 0$ or $1, j = 1, \cdots, n$, and c is a constant. $\phi(X_1, \cdots, X_n)$ is the probability of rejecting the null hypothesis H_0 under observation $X_1, \cdots, X_n$. We denote by ϕ_T the test ϕ defined by (1).

If $T(z)$ is a constant function of z, the test ϕ_T is trivial. Two rank statistics $T(z)$ and $T'(z)$ define the same rank test if

$$T'(z) = \lambda T(z) + \beta \qquad \text{for all} \quad z \tag{2}$$

with positive λ and arbitrary β.

In what follows, we are interested only in non-trivial rank tests, and two statistics T' and T satisfying (2) may be considered as identical.

The size α_{ϕ_T} of test ϕ_T is given by

$$\alpha_{\phi_T} = \sum_z \phi_T(z) P(z \mid F, F) \tag{3}$$

and the power function $\beta_T(F, G)$ may be expressed as

$$\beta_T(F, G) = \sum_z \phi_T(z) P(z \mid F, G), \tag{4}$$

where $P(z \mid F, G)$ represents the probability of $Z = z$ when F and G are true distributions, and the summation $\sum_z$ is over all $z = (z_1, \cdots, z_n)$ with $z_i = 0$ or 1 such that $\sum_{i=1}^n z_i = n_2$. $P(z \mid F, G)$ may be expressed as follows:

$$P(z \mid F, G) = n_1! \, n_2! \int \cdots \int_{-\infty < u_1 \leq \cdots \leq u_n < +\infty} \prod_{j=1}^{n} d[F(u_j)^{1-z_j} G(u_j)^{z_j}]. \tag{5}$$

Since F is assumed to be continuous, (5) may be written:

$$P(z \mid F, G) = P(z, H) = n_1! \, n_2! \int \cdots \int_{0 \leq t_1 \leq \cdots \leq t_n \leq 1} \prod_{j=1}^{n} d[t_j^{1-z_j} H(t_j)^{z_j}] \tag{6}$$

where

$$(7) \qquad H(t) = G[F^{-1}(t)], \qquad 0 \leqq t \leqq 1.$$

$H(t)$ is a cumulative distribution function on $[0, 1]$. We shall denote by Ω the set of all possible H's associated with any given two-sample problem, i.e.,

$$(8) \qquad \Omega = \{H; H = GF^{-1}, (F, G) \varepsilon H_0 \text{ or } H_1\}.$$

The sets Ω corresponding to the two-sample problems mentioned in Section 2 are as follows:

PROBLEM (I): Ω_1 is the set of all cumulative distribution functions H over $[0, 1]$.

PROBLEM (II): Ω_2 is the set of all cumulative distribution functions H over $[0, 1]$ such that $H(t) \leqq t$ for all $0 \leqq t \leqq 1$.

PROBLEM (III): Ω_3 is the set of all symmetric cumulative distribution functions H over $[0, 1]$:

$$H(t) + H(1 - t) = 1, \qquad \text{for all } 0 \leqq t \leqq 1.$$

PROBLEM (IV): Ω_4 is the set of all cumulative distributions H over $[0, 1]$ with mean not smaller than $\frac{1}{2}$:

$$\int t\, dH(t) \geqq \tfrac{1}{2}.$$

The set Ω in any problem of the above type has the following properties: First, Ω is a convex set; i.e., $H_1, H_2 \varepsilon \Omega$, and $0 \leqq \lambda \leqq 1$ imply $\lambda H_1 + (1 - \lambda)H_2 \varepsilon \Omega$. Secondly, for a sequence $H_1, H_2, \cdots$ of distributions in Ω, the condition that

$$\int_0^1 f(t)\, dH(t) = \lim_{\nu \to \infty} \int_0^1 f(t)\, dH_\nu(t)$$

for any continuous function $f(t)$ on $[0, 1]$, implies that the distribution H also belongs to the set Ω. This last property is sometimes stated that the set Ω is closed in the weak* topology.[3]

4. Locally Most Powerful Rank Tests. A set of cumulative distribution functions $\{(F(x, \theta), G(x, \theta)) : 0 \leqq \theta \leqq \bar{\theta}\}$, where $\bar{\theta} > 0$, is called a *one-parameter family of alternatives* if the following conditions are satisfied:

(a) $(F(x, \theta), G(x, \theta)) \varepsilon H_1$, for $0 < \theta \leqq \bar{\theta}$,

(b) $F(x, 0) = G(x, 0)$,

and

(c) $H(t, \theta) = G[F^{-1}(t, \theta), \theta]$ is uniformly differentiable with respect to θ at $\theta = 0$. Here $H(t, \theta)$ is called uniformly differentiable at $\theta = 0$ if the convergence, as θ tends to 0, of $[H(t, \theta) - H(t, 0)]/\theta$ to $[\partial H(t, \theta)/\partial\theta]_{\theta=0}$ is uniform with respect to t.

A rank test ϕ_T is said to be *locally most powerful* if there exists a one-parameter family of alternatives $(F(x, \theta), G(x, \theta))$ such that ϕ_T is most powerful against

[3] Cf., e.g., Bourbaki [1].

the alternatives $(F(x, \theta), G(x, \theta))$, $0 < \theta < \theta_0$, for some *positive* number θ_0, i.e.,

$$\beta_{\phi_T}(F(x, \theta), G(x, \theta)) \geqq \beta_\phi(F(x, \theta), G(x, \theta)), \qquad 0 < \theta < \theta_0, \tag{9}$$

for any rank test ϕ with size α.

The calculation of locally most powerful rank tests will be done by the following theorem:

THEOREM 1: *The locally most powerful rank test ϕ_T against a one-parameter family $(F(x, \theta), G(x, \theta))$ is determined uniquely for each size of test defined by*

$$T(z) = \sum_{j=1}^{n} a_j z_j, \tag{10}$$

where

$$a_j = \binom{n-1}{j-1} \int_0^1 t^{j-1}(1-t)^{n-j}\, dQ(t), \qquad j = 1, \cdots, n, \tag{11}$$

$$Q(t) = \left[\frac{\partial H(t, \theta)}{\partial \theta}\right]_{\theta=0}, \tag{12}$$

$$H(t, \theta) = G[F^{-1}(t, \theta), \theta]. \tag{13}$$

PROOF: Let us define $P(z, \theta)$ by

$$P(z, \theta) = P(z \mid F(\cdot, \theta), G(\cdot, \theta)). \tag{14}$$

Then, by the Neyman-Pearson Lemma, any rank test ϕ is locally most powerful against $(F(\cdot, \theta), G(\cdot, \theta))$ if, and only if, ϕ is defined by the statistic

$$T(z) = \left[\frac{\partial P(x, \theta)}{\partial \theta}\right]_{\theta=0}. \tag{15}$$

On the other hand, differentiating (6) with respect to θ, and noting that $H(t, \theta)$ is uniformly differentiable at $\theta = 0$, we have

$$\left[\frac{\partial P(x, \theta)}{\partial \theta}\right]_{\theta=0} = n_1!\, n_2! \underset{0 \leq t_1 \leq \cdots \leq t_n \leq 1}{\int \cdots \int} \sum_{j=1}^{n} z_j\, dt_1 \cdots dt_{j-1}\, dQ_j(t)\, dt_{j+1} \cdots dt_n \tag{16}$$

where $Q(t)$ is defined by (12). Since

$$\underset{0 \leq t_1 \leq \cdots \leq t_{j-1} \leq t_j}{\int \cdots \int} dt_1 \cdots dt_{j-1} = \frac{1}{(j-1)!}\, t_j^{j-1}$$

and

$$\underset{t_j \leq t_{j+1} \leq \cdots \leq t_n \leq 1}{\int \cdots \int} dt_{j+1} \cdots dt_n = \frac{1}{(n-j)!}\, (1 - t_j)^{n-j}$$

the integral in (16) may be further simplified and we have

$$\begin{aligned} &\left[\frac{dP(z, \theta)}{d\theta}\right]_{\theta=0} \\ &\quad = n_1!\, n_2! \sum_{j=1}^{n} z_j [1/(j-1)!(n-j)!] \int_0^1 t^{j-1}(1-t)^{n-j}\, dQ(t). \end{aligned} \tag{17}$$

The relation (17), together with (15), proves the theorem. Q.E.D.

Theorem 1 easily implies the following

COROLLARY: *Any locally most powerful rank test is admissible.*

In what follows, we shall first investigate locally most powerful rank tests for a general two-sample problem in which the set Ω is convex and closed in the weak* topology, and obtain a criterion for a rank test to be locally most powerful. We then consider the class of all locally most powerful rank tests for various two-sample problems mentioned above.

5. The Set A: A Special Class of Locally Most Powerful Rank Tests. Let us now consider a general two-sample problem for which the set Ω of all corresponding H functions is convex and closed in the weak* topology. In this section we shall introduce the class of rank tests which are locally most powerful with respect to a special class of one-parameter families of alternatives.

Let H be an arbitrary distribution function in Ω, and consider a one-parameter family $(F(x, \theta), G(x, \theta))$, $0 \leqq \theta \leqq 1$, satisfying the condition that[4]

$$H(t, \theta) = (1 - \theta)t + \theta H(t), \qquad 0 \leqq t \leqq 1, \quad 0 \leqq \theta \leqq 1 \tag{18}$$

where

$$H(t, \theta) = G[F^{-1}(t, \theta), \theta].$$

By Theorem 1, a rank test ϕ_T is locally most powerful against the one-parameter family satisfying (18) if, and only if,

$$T(z) = \lambda \sum_{j=1}^{n} a_j(H) z_j + \beta, \tag{19}$$

where λ is a positive number, β an arbitrary number, and

$$a_j(H) = \binom{n-1}{j-1} \int_0^1 t^{j-1}(1-t)^{n-j}\, dH(t), \qquad j = 1, \cdots, n. \tag{20}$$

We shall define the set A of n-dimensional vectors by

$$A = \{a = (a_1, \cdots, a_n);\ a_j = \lambda a_j(H) + \beta, j = 1, \cdots, n, H \varepsilon \Omega, \lambda \geqq 0, \text{ and } \beta \text{ is an arbitrary number}\}. \tag{21}$$

The set A, in other words, consists of all vectors that describe rank tests locally most powerful with respect to one-parameter families satisfying (18). It may be noted in particular that

$$a_j(H_0) = \frac{1}{n}, \qquad j = 0, 1, \cdots, n-1, \tag{22}$$

where

$$H_0(t) = t, 0 \leqq t \leqq 1.$$

[4] The case in which $H(t)$ is a polynomial was considered by Lehmann [5].

It will first be seen that the set A is a closed convex set. For any $H_1, H_2 \varepsilon \Omega$, and $0 \leqq \lambda \leqq 1$, we have $a[(1-\lambda)H_1 + \lambda H_2] = (1-\lambda)a(H_1) + \lambda a(H_2)$, which, together with the definition (21) of A, implies that the set A is a convex cone.

In order to prove the closedness of the set A, let $\{(a_1^\nu, \cdots, a_n^\nu)\}$ be a sequence of n-vectors in A which converges to an n-vector $a^0 = (a_1^0, \cdots, a_n^0)$. Since a^ν is in A, there exist $H^\nu \varepsilon \Omega$, $\lambda^\nu \geqq 0$, and β^ν such that

$$a_j^\nu = \lambda^\nu a_j(H^\nu) + \beta^\nu, j = 1, \cdots, n, \quad \nu = 1, 2, \cdots. \tag{23}$$

Taking a suitable subsequence of $\{H^\nu\}$, if necessary, we may without loss of generality suppose[5] that for any continuous function $f(t)$ on [0, 1],

$$\lim_{\nu\to\infty} \int_0^1 f(t)\, dH^\nu(t) = \int_0^1 f(t)\, dH(t), \tag{24}$$

with some distribution function $H(t)$ over [0, 1]. Since Ω is closed in the weak* topology, the function H belongs to the set Ω.

If the sequences $\{\lambda^\nu\}$ and $\{\beta^\nu\}$ are bounded, then, for any limiting points λ^0 and β^0, we have, by (23) and (24), $a_j^0 = \lambda^0 a_j(H) + \beta^0, j = 1, \cdots, n$, which shows that vector a belongs to the set A. If both sequences $\{\lambda^\nu\}$ and $\{\beta^\nu\}$ are unbounded, then we have $H(t) = t$, for all t. Hence, vector a trivially belongs to the set A.

6. The Set B. In order to investigate further the structure of the set A, we now introduce linear transformation L which maps n-vector $a = (a_1, \cdots, a_n)$ to b-vector $b = (b_0, b_1, \cdots, b_{n-1})$ defined by

$$L(a) = (L_0(a), L_1(a), \cdots, L_{n-1}(a)),$$

where

$$L_j(a) = \sum_{s=j+1}^{n} \binom{n-j-1}{s-j-1} \Big/ \left[a_s \Big/ \binom{n-1}{s-1} \right], \quad j = 0, 1, \cdots, n-1. \tag{25}$$

The inverse linear transformation L^{-1} of L is defined by

$$L^{-1}(b) = (L_1^{-1}(b), \cdots, L_n^{-1}(b)),$$

where

$$L_j^{-1}(b) = \sum_{s=j-1}^{n-1} (-1)^{s-j+1} \binom{n-j}{s-j+1} b_s, \quad j = 1, \cdots, n. \tag{26}$$

The fact that the linear transformation L^{-1} defined by (26) is the inverse of L defined by (25) is easily seen from the following identities:

$$t^j = \sum_{s=j+1}^{n} \binom{n-j-1}{s-j-1} t^{s-1}(1-t)^{n-s}, \quad j = 0, 1, \cdots, n-1, \tag{27}$$

and

$$t^{j-1}(1-t)^{n-j} = \sum_{s=j-1}^{n-1} (-1)^{s-j+1} \binom{n-j}{s-j+1} t^s, \quad j = 1, \cdots, n. \tag{28}$$

[5] Cf., e.g., Bourbaki [2].

Let us now define the set B as the image of the set A by the linear transformation L:

$$B = \{b = (b_0, \cdots, b_{n-1}) : b = L(a) \quad \text{for some} \quad a \in A\}. \tag{29}$$

Since the set A is a closed convex cone and L is linear, the set B again is a closed convex cone in the n-vector space E^n. It is noted that we have, by the identities (27) and (28),

$$L_j(a(H)) = \int_0^1 t^j \, dH(t), \qquad j = 0, 1, \cdots, n-1, \tag{30}$$

for any $H \in \Omega$. We have, in particular, that

$$L_j(1, \cdots, 1) = n/(j+1), \qquad j = 0, 1, \cdots, n-1. \tag{31}$$

The definitions (21) and (29) of the sets A and B, together with (30) and (31), imply that an n-vector $b = (b_0, \cdots, b_{n-1})$ belongs to the set B if, and only if, there exist $H \in \Omega$, $\lambda \geqq 0$ and real number β such that

$$b_j = \lambda \int_0^1 t^j \, dH(t) + \beta/(j+1), \qquad j = 0, 1, \cdots, n-1. \tag{32}$$

We shall give a necessary and sufficient condition for an n-vector $b = (b_0, \cdots, b_{n-1})$ to be in the set B.

We first define the *polar cone* B^* of any set B of n-vectors $b = (b_0, \cdots, b_{n-1})$ as the set of all n-vectors $y = (y_0, \cdots, y_{n-1})$ whose inner product with any vector in B is non-negative:

$$B^* = \{y = (y_0, \cdots, y_{n-1}) : y \cdot b \geqq 0 \quad \text{for all} \quad b \in B\}, \tag{33}$$

where $y \cdot b$ denotes the inner product of two vectors y and b:

$$y \cdot b = \sum_{j=0}^{n-1} y_j b_j .$$

For any n-vector $y = (y_0, \cdots, y_{n-1})$, let us define the polynomial $y(t)$ by

$$y(t) = \sum_{j=0}^{n-1} y_j t^j. \tag{34}$$

By the definition (33), and the relation (32), an n vector $y = (y_0, \cdots, y_{n-1})$ belongs to the set B^* if, and only if,

$$\sum_{j=0}^{n-1} y_j \left[\lambda \int_0^1 t^j \, dH(t) + \beta/(j+1) \right] \geqq 0 \tag{35}$$

for all $H \in \Omega$, $\lambda \geqq 0$, and β real.

The relation (35), in view of (34), is equivalent to the following:

$$\int_0^1 y(t) \, dH(t) \geqq 0, \qquad \text{for all } H \in \Omega, \text{and } \int_0^1 y(t) \, dt = 0.$$

Therefore, we have

LEMMA 1. *An n-vector* $y = (y_0, \cdots, y_{n-1})$ *belongs to the set* B^* *if, and only if,*

$$\int_0^1 y(t)\, dH(t) \geqq 0, \qquad \text{for all} \quad H \varepsilon \Omega, \tag{36}$$

and

$$\int_0^1 y(t)\, dt = 0. \tag{37}$$

On the other hand, since the set B is a closed convex cone in the n-vector space E^n, we have, by the duality theorem[6] on closed convex cones, that

$$B^{**} = B. \tag{38}$$

The relation (38) may be expressed as

LEMMA 2. *An n-vector* $b = (b_0, \cdots, b_{n-1})$ *belongs to the set B if, and only if,*

$$b \cdot y \geqq 0, \qquad \text{for all} \quad y \varepsilon B^*.$$

7. The Two-Sided Two-Sample Problem. We shall first consider Problem (I): Test the null hypothesis $H_0 : F = G$ against the alternatives that $H_1 : F \neq G$. In this case, the set Ω_1 consists of all cumulative distribution functions H over [0, 1].

The class of all locally most powerful rank tests is characterized by the following theorem:

THEOREM 2: *A non-trivial rank test* ϕ_T *is locally most powerful for Problem* (I) *if, and only if,*

$$T(z) = \sum_{j=1}^{n} a_j z_j, \tag{40}$$

where $a_1 \cdots, a_n$ *are arbitrary constants.*

PROOF: It will be shown first that the set B^* consists of the zero vector $0 = (0, \cdots, 0)$ alone. Indeed, let an n-vector $y = (y_0, \cdots, y_{n-1})$ belong to B^*. By Lemma 1, the conditions (36) and (37) must be satisfied, where Ω_1 is the set of all cumulative distribution functions H on [0, 1]. Then the condition (36) implies that

$$y(t) \geqq 0 \quad \text{for all} \quad 0 \leqq t \leqq 1. \tag{41}$$

But since $y(t)$ is a polynomial, the relation (36), together with (37), implies that $y(t) = 0$, for all $0 \leqq t \leqq 1$. Hence, $y_j = 0, j = 0, 1, \cdots, n - 1$.

The polar cone B^{**} of $B^* = (0)$ is now the set E^n of all n-vectors. By Lemma 2, therefore, we have $B = E^n$. Hence, the set A also is equal to the set E^n of all n-vectors, and any test ϕ_T defined in terms of $T(z)$ of the form (40) is locally most powerful.

---

[6] Cf., e.g., Bourbaki [1] or Fenchel [3].

8. The One-Sided Two Sample Problem. In this section we will be concerned with Problem (II): Test the hypothesis $H_0 : F = G$ against the alternatives that $H_2 : F \geq G$. The space Ω_2 for this problem consists of all cumulative distribution functions H on [0, 1] such that

$$H(t) \leqq t, \quad 0 \leqq t \leqq 1. \tag{42}$$

Before stating the characterization of the class of all locally most powerful rank tests for Problem (II), we introduce some concepts from the Hausdorff theory of moments.[7]

An m-vector $c = (c_0, c_1, \cdots, c_{m-1})$ may be called here a solution to the m-dimensional moment problem over [0, 1] if there exist a distribution function H over [0, 1] and a non-negative number λ such that

$$c_j = \lambda \int_0^1 t^j \, dH(t), \qquad j = 0, 1, \cdots, m-1. \tag{43}$$

Let C_m be the set of all solutions to the m-dimensional moment problem over [0, 1]:

$$C_m = \Big\{ c = (c_0, c_1, \cdots, c_{m-1}) : c_j = \lambda \int t^j \, dH(t), \; j = 0, 1, \cdots, m-1, \text{for some distribution } H \text{ over } [0,1] \text{ and non-negative number } \lambda \Big\}. \tag{44}$$

Similar to the set B, the set C_m here is also a closed convex cone in the m-vector space.

The polar cone C_m^* may be written as

$$\begin{aligned} C_m^* &= \left\{ z = (z_0, \cdots, z_{m-1}) : \int_0^1 z(t) \, dH(t) \geqq 0 \text{ for all distributions } H \right\} \\ &= \{ z = (z_0, \cdots, z_{m-1}) : z(t) \geqq 0, \text{ for all } 0 \leqq t \leqq 1 \}, \end{aligned} \tag{45}$$

where $z(t)$ is defined by $z(t) = \sum_{j=0}^{m-1} z_j t^j$. Hence, by the duality theorem on closed convex cones, we have that

LEMMA 3. *For an m-vector $c = (c_0, \cdots, c_{m-1})$, $c \in C_m$ if, and only if, $c \cdot z \geqq 0$, for all $z = (z_0, \cdots, z_{m-1})$ such that $z(t) \geqq 0$, $0 \leqq t \leqq 1$.*

By a theorem from the Hausdorff theory of moment problems,[8] we have, on the other hand, that an m-vector $c = (c_0, \cdots, c_{m-1})$ is a solution to the m-dimensional moment problem over [0, 1] if, and only if, the Hankel determinants $\Delta_s(c_0, \cdots, c_s)$ and $\bar{\Delta}_s(c_0, \cdots, c_s)$ are all non-negative, for all $s = 0, 1, \cdots, m-1$.

[7] For the Hausdorff theory of moments, the reader is referred to, e.g., Shohat and Tamarkin [7] or Karlin and Shapley [4].

[8] Cf. Karlin and Shapley [4], pp. 54–57.

The *Hankel determinants* $\Delta_s(c_0, \cdots, c_s)$ and $\bar{\Delta}_s(c_0, \cdots, c_s)$ are defined by

$$\Delta_{2r}(c_0, \cdots, c_{2r}) = \begin{vmatrix} c_0 & c_1 & \cdots & c_r \\ \vdots & & & \\ c_r & c_{r+1} & \cdots & c_{2r} \end{vmatrix}$$

$$\Delta_{2r+1}(c_0, \cdots, c_{2r+1}) = \begin{vmatrix} c_1 & c_2 & \cdots & c_{r+1} \\ \vdots & & & \\ c_{r+1} & c_{r+2} & \cdots & c_{2r+1} \end{vmatrix}$$

$$\bar{\Delta}_{2r}(c_0, \cdots, c_{2r}) = \begin{vmatrix} c_1 - c_2 & c_2 - c_3 & \cdots & c_r - c_{r+1} \\ \vdots & & & \\ c_r - c_{r+1} & c_{r+1} - c_{r+2} & \cdots & c_{2r-1} - c_{2r} \end{vmatrix}$$

$$\bar{\Delta}_{2r+1}(c_0, \cdots, c_{2r+1}) = \begin{vmatrix} c_0 - c_1 & c_1 - c_2 & \cdots & c_r - c_{r+1} \\ \vdots & & & \\ c_r - c_{r+1} & c_{r+1} - c_{r+2} & \cdots & c_{2r} - c_{2r+1} \end{vmatrix}$$

The class of all locally most powerful rank tests for Problem (II) may now be characterized by the following:

THEOREM 3. *A non-trivial rank test ϕ_T is locally most powerful for Problem* (II) *if, and only if,* $T(z_1, \cdots, z_n) = \sum_{j=1}^{n} a_j z_j$, *and* $(c_0, c_1, \cdots, c_{n-2})$ *has all non-negative Hankel determinants, where*

$$c_j = \left[1 \Big/ \binom{n-2}{j}\right] \sum_{s=j}^{n-2} \binom{s+1}{j+1} (a_{s+2} - \bar{a}), \quad j = 0, \cdots, n-2, \tag{46}$$
$$\bar{a} = \frac{1}{n} \sum_{s=1}^{n} a_s .$$

PROOF: In the present case, the set Ω_2 consists of all cumulative distribution functions $H(t)$ over $[0, 1]$ such that

$$H(t) \leq t, \quad \text{for all} \quad 0 \leq t \leq 1. \tag{47}$$

By Lemma 1, an $\bar{n}$-vector $y = (y_0, \cdots, y_{n-1})$ belongs to the polar cone B^* if, and only if,

$$\int_0^1 y(t)\, dH(t) \geq 0, \text{ for all } H \text{ such that } H(t) \leq t, 0 \leq t \leq 1, \tag{48}$$

and

$$\int_0^1 y(t)\, dt = 0. \tag{49}$$

We shall show that in view of (49) condition (48) may be replaced by

$$y'(t) \geq 0 \quad \text{for all} \quad 0 \leq t \leq 1. \tag{50}$$

In fact, for any polynomial $y(t)$ satisfying (49), we have, by a partial integra-

tion, that

$$\begin{aligned}\int_0^1 y(t)\,dH(t) &= [y(t)H(t)]_0^1 - \int_0^1 y'(t)H(t)\,dt \\ &= \int_0^1 y'(t)(t - H(t))\,dt.\end{aligned} \tag{51}$$

Now suppose that there exists t_0 such that $y'(t_0) < 0, 0 \leqq t_0 \leqq 1$. Then, by the continuity of $y'(t)$, there is an interval I in [0, 1] containing t_0 such that $y'(t) < 0$, for all $t \varepsilon I$. It is then possible to construct a cumulative distribution function H_1 over [0, 1] such that

$$\begin{cases} t - H_1(t) = 0, & \text{for all } t \varepsilon I, \\ t - H_1(t) > 0, & \text{for } t \text{ interior to } I. \end{cases}$$

Then H_1 belongs to the set Ω, and by (50) and (51),

$$\int_0^1 y'(t)\,dH_1(t) = \int_I y'(t)\,dH_1(t) < 0,$$

which contradicts (48) and (51). Therefore, we have $y'(t) \geqq 0$, for all $0 \leqq t \leqq 1$.

The polar cone B^*, therefore, may be characterized by

$$B^* = \left\{ y = (y_0, \cdots, y_{n-1}) : y'(t) \geqq 0, \text{ for all } 0 \leqq t \leqq 1, \text{ and } \int_0^1 y(t)\,dt = 0 \right\}. \tag{52}$$

We may rewrite (52) as follows: $y \varepsilon B^*$ if, and only if,

$$y_j = \frac{1}{j} z_{j-1}, \qquad j = 1, \cdots, n-2, \tag{53}$$

and

$$y_0 = -\sum_{j=1}^{n-1} \frac{1}{j+1} y_j, \tag{54}$$

for some $z = (z_0, \cdots, z_{n-2}) \varepsilon C^*_{n-1}$.

By Lemma 2, (53) and (54) imply that $b = (b_0, b_1, \cdots, b_{n-1}) \varepsilon B$ if, and only if,

$$\sum_{j=1}^{n-1} \frac{1}{j} \left(b_j - \frac{1}{j+1} b_0 \right) z_{j-1} \geqq 0 \quad \text{for all} \quad z = (z_0, \cdots, z_{n-2}) \varepsilon C^*_{n-1}. \tag{55}$$

By Lemma 3, the relation (55) is satisfied if, and only if,

$$c_j = \frac{1}{j+1} \left(b_{j+1} - \frac{1}{j+2} b_0 \right), \quad j = 0, 1, \cdots, n-2, \tag{56}$$

have all non-negative Hankel determinants. Substituting (25) into (56), c_j are expressed by (46).

We now show that any locally most powerful rank test ϕ_T may be expressed in terms of $T(z) = \sum_{j=1}^n a_j z_j$ with $a = (a_1, \cdots, a_n) \varepsilon A$. In fact, let ϕ_T be locally most powerful against a one-parameter family $(F(x, \theta), G(x, \theta))$. By Theorem 1, we have $T(z) = \sum_{j=1}^n a_j z_j$, where a_j are defined by (11). Since

$$H(0, \theta) = 0, \quad H(1, \theta) = 1,$$

$$H(t, \theta) \leqq t = H(t, 0), \quad 0 \leqq t \leqq 1,$$

we have

$$Q(0) = Q(1) = 0, \tag{57}$$

$$Q(t) \leqq 0, \quad 0 \leqq t \leqq 1, \tag{58}$$

where $Q(t)$ is defined by (12).

Let us first consider the case where $Q(t)$ is *continuously differentiable* on [0, 1]. Then, by (57) and (58), there exists a *positive* number λ such that $H_1(t) = t + \lambda Q(t)$ is a cumulative distribution function over [0, 1], for which we have

$$H_1(t) \leqq t, \quad 0 \leqq t \leqq 1. \tag{59}$$

Consider a one-parameter family $(F_1(x, \theta), G(x, \theta))$ satisfying

$$H_1(t, \theta) = (1 - \theta)t + \theta H_1(t). \tag{60}$$

$H_1(t, \theta)$ belongs to the set Ω_2 for $0 \leqq \theta \leqq 1$, and

$$\left[\frac{\partial H_1(t, \theta)}{\partial \theta}\right]_{\theta=0} = \lambda Q(t). \tag{61}$$

Therefore, $\sum_j a_j z_j$ is locally most powerful against the one-parameter family $(F_1(x, \theta), G_1(x, \theta))$ satisfying (60), and $a = (a_1, \cdots, a_n)$ belongs to the set A for Problem (II).

Now consider the general case where $Q(t)$ is not necessarily differentiable. Let $\{H_\nu(t, \theta); \nu = 1, 2, \cdots\}$ be a sequence of cumulative distribution functions in Ω such that

$$\lim_{\nu\to\infty} Q_\nu(t) = Q(t), \qquad 0 \leqq t \leqq 1, \tag{62}$$

where

$$Q_\nu(t) = \left[\frac{\partial H_\nu(t, \theta)}{\partial \theta}\right]_{\theta=0} \tag{63}$$

is continuously differentiable with respect to t. The locally most powerful rank order test against one-parameter family $(1 - \theta)t + \theta H_\nu(t)$ is defined by

$$T^\nu(z) = \sum_{j=1}^n a_j^\nu z_j \tag{64}$$

where

$$a_j^\nu = \binom{n-1}{j-1} \int_0^1 t^{j-1}(1 - t)^{n-j}\, dQ_\nu(t). \tag{65}$$

Then the relations (62) and (65) imply that

$$\lim_{r\to\infty} a_j^r = a_j, \qquad j = 1, \cdots, b. \tag{66}$$

The vector $a^r = (a_1^r, \cdots, a_n^r)$ belongs to A which is a closed set. Hence, by (66), we have $a = (a_1, \cdots, a_n) \varepsilon A$.

9. The Symmetric Two-Sided Two-Sample Problem. In this section we investigate the structure of the class of all locally most powerful rank tests for Problem (III). Problem (III) is to test the null hypothesis $H_0 : F = G$, symmetric against the alternatives that $H_3 : F \neq G$, symmetric with the same median. For Problem (III), the set Ω_3 consists of all cumulative distributions H over [0, 1] such that

$$H(t) + H(1 - t) = 1, \qquad \text{for all} \quad 0 \leqq t \leqq 1. \tag{67}$$

The class of all locally most powerful rank tests is characterized by the following

THEOREM 4. *A non-trivial rank test ϕ_T is locally most powerful for Problem* (III) *if, and only if,* $T(z_1, \cdots, z_n) = \sum_{j=1}^{n} a_j z_j$ *with*

$$\sum_{s=j+1}^{n} \binom{s-1}{j} (a_{n+1-s} - a_s) = 0, \qquad \textit{for all} \quad j = 1, \cdots, n. \tag{68}$$

PROOF: Since the set Ω_3 consists of all cumulative distribution functions H satisfying (67), Lemma 1 implies that an n-vector $y = (y_0, \cdots, y_{n-1})$ belongs to the polar cone B^* for Problem (III) if, and only if,

$$\int_0^1 y(t)\, dH(t) \geqq 0, \qquad \text{for all cumulative distribution functions } H(t) \text{ satisfying (67)}, \tag{69}$$

and

$$\int_0^1 y(t)\, dt = 0. \tag{70}$$

If H satisfies (67), we have

$$\int_0^1 y(t)\, dH(t) = \int_0^{1/2} [y(t) + y(1 - t)]\, dH(t).$$

Relations (69) and (70) may now be replaced by

$$\int_0^{1/2} [y(t) + y(1 - t)]\, dH(t) \geqq 0, \qquad \text{for any cumulative distribution functions } H \text{ over } [0, \tfrac{1}{2}], \tag{71}$$

and

$$\int_0^{1/2} [y(t) + y(1 - t)]\, dt = 0. \tag{72}$$

But relations (71) and (72) are satisfied if, and only if, $y(t) + y(1 - t) = 0$,

$0 \leq t \leq \frac{1}{2}$. Hence, an n-vector $y = (y_0, \cdots, y_{n-1})$ belongs to the polar cone B^* if, and only if,

$$y(t) + y(1 - t) = 0 \qquad \text{for all} \quad 0 \leq t \leq 1. \tag{73}$$

The relation (73) may be written

$$y_j + \sum_{s=j}^{n-1} (-1)^j \binom{s}{j} y_s = 0, \qquad j = 0, 1, \cdots, n-1. \tag{74}$$

By Lemma 2, the set B is equal to the polar cone B^{**} of B^*. Therefore, by (74), an n-vector $b = (b_0, \cdots, b_{n-1})$ is in B if, and only if,

$$b_j = u_j + \sum_{s=0}^{j} (-1)^s \binom{s}{j} u_s, \qquad j = 0, 1, \cdots, n-1, \tag{75}$$

for some $u = (u_0, \cdots, u_{n-1})$.

It may be noted that, for an n-vector $b = (b_0, \cdots, b_{n-1})$, there exists an n-vector $u = (u_0, \cdots, u_{n-1})$ for which the relations (75) are satisfied if, and only if,

$$b_j = \sum_{s=0}^{j} (-1)^s \binom{j}{s} b_s, \qquad j = 0, 1, \cdots, n-1. \tag{76}$$

It is evident that the relation (75) implies (76). On the other hand, let the relation (76) be satisfied. In order to prove the existence of $u_0, \cdots, u_{n-1}$ satisfying (75), we use the mathematical induction on n. Let us assume that we have found $u_0, \cdots, u_{n-2}$ which satisfy the relations (75) for $j = 0, 1, \cdots, n-2$. If $n - 1$ is an even number, u_{n-1} may be determined by

$$2u_{n-1} = b_{n-1} - \sum_{s=0}^{n-2} (-1)^s \binom{n-1}{s} u_s.$$

$u_0, \cdots, u_{n-2}$ and u_{n-1} then satisfy (75) for $s = 0, \cdots, n-1$.

If $n - 1$ is an odd number, the relation (75) may be satisfied with $u_0, \cdots, u_{n-2}$, and arbitrary u_{n-1}. In fact, by (76),

$$b_{n-1} = \sum_{s=0}^{n-1} (-1)^s \binom{n-1}{s} b_s = -b_{n-1} + \sum_{s=0}^{n-2} (-1)^s \binom{n-1}{s} b_s.$$

Hence,

$$\begin{aligned} 2b_{n-1} &= \sum_{s=0}^{n-2} (-1)^s \binom{n-1}{s} \left[u_s + \sum_{r=0}^{s} (-1)^r \binom{s}{r} u_r \right] \\ &= \sum_{s=0}^{n-2} (-1)^s \binom{n-1}{s} u_s + \sum_{r=0}^{n-2} \sum_{s=r}^{n-2} (-1)^{r+s} \binom{n-1}{s} \binom{s}{r} u_r. \end{aligned} \tag{77}$$

But

$$\sum_{s=r}^{n-2} (-1)^{r+s} \binom{n-1}{s} \binom{s}{r} = \binom{n-1}{r} \sum_{k=0}^{n-r-2} (-1)^k \binom{n-r-1}{k} = -(-1)^{n-r-1} \binom{n-1}{r}.$$

Since $n - 1$ is odd, we may now write (77) as follows:

$$(78) \qquad 2b_{n-1} = 2\sum_{r=0}^{n-2} (-1)^r \binom{n-1}{r} u_r .$$

Dividing (78) by 2 implies the relation (75) for $j = n - 1$. Expressing (76) in terms of $a_1, \cdots, a_n$, we get the relation (68).

We now have to show that the set A actually exhausts the class of all locally most powerful rank tests for Problem (III). Let $T(z) = \sum_{j=1}^n a_j z_j$ define the rank test which is locally most powerful against one-parameter family $(F(x, \theta), G(x, \theta))$ such that

$$(79) \qquad H(t, \theta) + H(1 - t, \theta) = 1, \qquad 0 \leqq t \leqq 1.$$

Since the set A is a closed set, it again suffices to consider the case in which $Q(t) = [\partial H(t, \theta)/\partial\theta]_{\theta=0}$ is continuously differentiable.

By (79) we have

$$(80) \qquad Q(0) = Q(1) = 0, \qquad Q(t) + Q(1 - t) = 0, \qquad 0 \leqq t \leqq 1.$$

Since $Q(t)$ is continuously differentiable, there exists a positive number λ such that $H_1(t) = t + \lambda Q(t)$ is a symmetric cumulative distribution function over $[0, 1]$. Hence, for some constant β, we have $a_j = \lambda a_j(H_1) + \beta$, which shows that $a = (a_1, \cdots, a_n)$ belongs to the set A for Problem (III).

10. The case $\int_0^1 F\, dG \geqq \frac{1}{2}$. We shall finally consider Problem (IV): Test the null hypothesis $H_0 : F = G$ against the alternatives that $H_4 : F \neq G, \int_0^1 F\, dG \geqq \frac{1}{2}$. The set Ω_4 for Problem (IV) is the set of all cumulative distributions H over $[0, 1]$ such that

$$(81) \qquad \int_0^1 t\, dH(t) \geqq \tfrac{1}{2}.$$

THEOREM 5. *A non-trivial rank test ϕ_T is locally most powerful for Problem* (IV) *if, and only if,* $T(z) = \sum_{j=1}^n a_j z_j$ *with*

$$(82) \qquad \sum_{s=1}^{[(n+1)/2]} ((n + 1)/2 - s)(a_{n+1-s} - a_s) \geqq 0,$$

where $[(n + 1)/2]$ *denote the greatest integer less than or equal to* $(n + 1)/2$.

PROOF: The polar cone B^* in the present case consists of all n-vectors $y = (y_0, \cdots, y_{n-1})$ such that

$$(83) \qquad \int_0^1 y(t)\, dH(t) \geqq 0, \qquad \text{for all } H \text{ for which} \qquad \int_0^1 t\, dH(t) \geqq \tfrac{1}{2},$$

and

$$(84) \qquad \int_0^1 y(t)\, dt = 0.$$

The set Ω_4 in particular contains the set of all cumulative distribution functions H such that $H(t) \leqq t$, for $0 \leqq t \leqq 1$. Therefore, by an argument similar to the

one in Problem (II), we have

$$y'(t) \geqq 0, \qquad 0 \leqq t \leqq 1, \qquad \text{for all} \quad y \in B^*. \tag{85}$$

Similarly, since the set Ω_4 contains all cumulative distribution functions H over $[0, 1]$ such that $H(t) + H(1 - t) = 1, 0 \leqq t \leqq 1$, we have that

$$y(t) + y(1 - t) = 0, \qquad 0 \leqq t \leqq 1, \qquad \text{for all} \quad y \in B^*. \tag{86}$$

We shall show furthermore that, for any $y \in B^*$,

$$y(t) \quad \text{is linear in} \quad t, \qquad 0 \leqq t \leqq 1. \tag{87}$$

In fact, let σ and τ be arbitrary numbers between 0 and $\frac{1}{2}$, and let $H_{\sigma,\tau}$ be the cumulative distribution function corresponding to the following probability distribution:

$$\text{Prob.}\ \{t = \tfrac{1}{2} - b\} = \tau/(\sigma + \tau),$$
$$\text{Prob.}\ \{t = \tfrac{1}{2} + \tau\} = \sigma/(\sigma + \tau).$$

The cumulative distribution function $H_{\sigma,\tau}$ belongs to the set Ω_4 for Problem (IV), and

$$\int_0^1 y(t)\, dH_{\sigma,\tau}(t) = \frac{\tau}{\sigma + \tau}\, y\left(\tfrac{1}{2} - \sigma\right) + \frac{\sigma}{\sigma + \tau}\, y\left(\tfrac{1}{2} + \tau\right).$$

Therefore, if $y = (y_0, \cdots, y_{n-1}) \in B^*$, the relation (83) implies that

$$-y(\tfrac{1}{2} - \sigma)/\sigma \leqq y(\tfrac{1}{2} + \tau)/\tau, \qquad \text{for all} \quad 0 < \sigma, \tau < \tfrac{1}{2}. \tag{88}$$

The relation (88), together with (86), implies that $y(t)$ be linear in $t, 0 \leqq t \leqq 1$. Therefore, if y belongs to the polar set B^*, we have, by (85), (86), and (87),

$$y_0 = -\tfrac{1}{2}\, y_1, \qquad y_1 \geqq 0, \qquad y_2 = \cdots = y_{n-1} = 0. \tag{89}$$

On the other hand, let an n-vector $y = (y_0, \cdots, y_{n-1})$ satisfy the relation (89). Then, for any $H \in \Omega_4$,

$$\int_0^1 y(t)\, dH(t) = y_0 + y_1 \int_0^1 t\, dH(t) \geqq y_0 + \tfrac{1}{2}\, y_1 = 0,$$

and

$$\int_0^1 y(t)\, dt = y_0 + \tfrac{1}{2}\, y_1 = 0.$$

The vector y, therefore, belongs to the polar cone B^*. Hence, the polar cone B^* consists of all n-vectors $y = (y_0, \cdots, y_{n-1})$ satisfying the relation (89). The set B is, by Lemma 2, equal to the polar cone B^{**} to B^*. Thus, we have

$$b = (b_0, \cdots, b_{n-1}) \in B \quad \text{if, and only if,} \quad b_1 \geqq \tfrac{1}{2}\, b_0. \tag{90}$$

Writing the relation (90) in terms of a's, we have that $a = (a_1, \cdots, a_n) \in A$ if, and only if, (82) is satisfied.

Let ϕ_T be locally most powerful against a one-parameter family $(F(x, \theta), G(x, \theta))$. By Theorem 1, we have $T(z) = \sum_{j=1}^n a_j z_j$, where a_j are defined by (11). Since

$$H(0) = 0, \qquad H(1) = 1, \qquad \int_0^1 t\, dH(t, 0) \geqq \tfrac{1}{2} = \int_0^1 t\, dH(t, 0), \qquad 0 \leqq t \leqq 1,$$

we have

$$Q(0) = Q(1) = 0, \qquad \int_0^1 t\, dQ(t) \geqq 0. \tag{91}$$

It again suffices to consider the case where $Q(t)$ is continuously differentiable at every point t. By (91) there exists a positive number λ such that $H_1(t) = t + \lambda Q(t)$ is a cumulative distribution function over [0, 1] and

$$\int t\, dH_1(t) = \tfrac{1}{2} + \lambda \int t\, dQ(t) \geqq \tfrac{1}{2}.$$

Consider a one-parameter family $(F_1(x, \theta), G(x, \theta))$ satisfying $H_1(t, \theta) = (1 - \theta)t + \theta H_1(t)$. $H_1(t, \theta)$ belongs to the set Ω_A for $0 \leqq \theta \leqq 1$, and, for some number β,

$$\left[\frac{\partial H_1(t, \theta)}{\partial \theta}\right]_{\theta=0} = \lambda Q(t) + \beta.$$

The vector $a = (a_1, \cdots, a_n)$, therefore, belongs to the set A for Problem (IV).

11. Acknowledgments. I received invaluable suggestions and comments from the members of the Summer Statistical Institute on Nonparametric Methods at the University of Minnesota in 1958, especially from W. Hoeffding, S. Karlin, I. R. Savage, M. Sobel, and C. K. Tsao. The earlier version of the manuscript contained some errors, particularly with respect to the application of the theory of moments, which were pointed out by S. Karlin, and the correct version owes much to discussions with him. I am also indebted to the referees for valuable comments. I wish to express my indebtedness and gratitude to all, although I am solely responsible for any shortcomings of the paper.

REFERENCES

[1] N. Bourbaki, *Espaces Vectoriels Topologiques*, Livre V, Première Partie, *Eléments de Mathématique*, Hermann, Paris, 1953 and 1955.

[2] N. Bourbaki, *Integration*, Livre VI, Première Partie, *Eléments de Mathématique*, Hermann, Paris, 1952 and 1956.

[3] W. Fenchel, *Convex Cones, Sets, and Functions*, Princeton University, Princeton, 1953.

[4] S. Karlin and L. S. Shapley, *Geometry of Moment Spaces*, Memoirs of the American Mathematical Society, No. 12, 1953.

[5] E. L. Lehmann, "The power of rank tests," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 23–43.

[6] I. R. Savage, "Contributions to the theory of rank order statistics–the two-sample case," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 590–615.

[7] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*, Mathematical Survey, No. 1, American Mathematical Society, 1943.

ON THE STRUCTURE OF DISTRIBUTION-FREE STATISTICS[1]

C. B. Bell[2]

Stanford University and San Diego State College

Introduction and Summary. Let $X_1, X_2, \cdots, X_n$ be a sample of a one-dimensional random variable X which has the continuous cumulative probability function (cpf) F. It has been observed that the distribution-free statistics commonly appearing in the literature can be written in the form $\Phi[F(X_1), F(X_2), \cdots, F(X_n)]$ where Φ is a measurable symmetric function defined on the unit cube. Such statistics are said to have structure (d).

Birnbaum and Rubin [12] have proved that for the family Ω^*, of strictly monotone continuous cpf's, statistics of structure (d) possess a property stronger than that of being distribution-free.

The purpose of this paper is to study the extension of the Birnbaum-Rubin (B-R) result to other classes of cpf's and to present a different approach to these results. It is found that a one-sided extension of the B-R result is valid for all properly closed, symmetrically complete classes of cpf's. Then, from the existing literature on completeness, one can conclude that the extension is valid for several other classes of statistical interest.

The relation between statistics of structure (d) and strongly distribution-free statistics (Section 1) is of importance for two reasons. First of all, if one is designing distribution-free tests, the results here and in [12] guarantee that if one chooses a statistic of structure (d), one has a strongly distribution-free statistic for several large classes of cpf's.

On the other hand if one has a strongly distribution-free statistic, the results guarantee that it is of structure (d). Hence, its cpf can be written as the volume of a polyhedral region in the n dimensional unit cube. Under such circumstances the work of Smirnov [20], Feller [13], Anderson and Darling [4], and Birnbaum [9] indicate that it should be possible to evaluate the cpf explicity; reduce it to a system of recursion formulae; tabulate it with the aid of high-speed computers or at least evaluate its limiting distribution.

This article is divided into four sections. In Section 1 distribution-free statistics of various types are introduced. Section 2 contains some preliminary results concerning cpf's. The main theorem is proved in Section 3; and Section 4 contains a survey of the known pertinent completeness results as well as a corollary of the main theorem.

1. Distribution-free Statistics. Consistent with the notation of Scheffé [18] and B-R [12] let

Ω_0 = the class of all cpf's;

Received December 26, 1957; revised July 27, 1959.

[1] This research was supported by National Science Foundation Grant NSFG-3662.

[2] Present address: San Diego State College, San Diego 15, California.

Ω_1 = the class of all non-degenerate cpf's;
Ω_2 = the class of all continuous cpf's;
Ω^* = the class of all strictly monotone continuous cpf's;
Ω_3 = the class of all absolutely continuous (with respect to Lebesgue measure) cpf's;
Ω_4 = the class of all cpf's with continuous derivatives;
Ω_u = the class of all cpf's which are uniform within intervals [11], [12]; and
Ω_e = the class of all cpf's with densities of the form

$$C(\theta_1, \cdots, \theta_n) \exp\{-x^{2n} - \theta_1 x - \theta_2 x^2 - \cdots - \theta_n x^n\},$$

[16]. Analogously, for the unit interval I, one defines
$\Omega_0(I)$ = the class of all cpf's on I;
$\Omega_1(I)$ = the class of all non-degenerate cpf's on I;
$\Omega_2(I)$ = the class of all continuous cpf's on I; etc.

If Ω and Ω' are two arbitrary families of cpf's, a real-valued function

$$S_G = S_G(X_1, X_2, \cdots, X_n)$$

will be called a statistic in Ω with regard to (w.r.t.) Ω', if for every $G \varepsilon \Omega$, and $F \varepsilon \Omega'$; and $X_1, X_2, \cdots, X_n$ in the n-dimensional sample space for a random variable X which has cpf F,

(a) $S_G(X^{(n)}) = S_G(X_1, X_2, \cdots, X_n)$ is defined everywhere in the sample space, and

(b) $S_G = S_G(X^{(n)})$ has a probability distribution; this probability distribution will be denoted by $\mathcal{P}_F^{(n)} S_G^{-1}$.

For example, consider von Mises' statistic

$$w_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - G(x)]^2 \, dG(x) = (1/12n) + \sum_{i=1}^{n} [G(X_i) - (2i-1)/n]^2;$$

Kolmogoroff's statistic

$$D_n = \sup_{-\infty<x<\infty} |F_n(x) - G(x)| = \max_{i=1,\cdots,n} [G(X_i') - (i-1)/n, (i/n) - G(X_i')];$$

Anderson and Darling's

$$K_n = \sup_{-\infty<x<\infty} \sqrt{n} |F_n(x) - G(x)| (\Psi[G(x)])^{\frac{1}{2}}$$

and

$$W_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - G(x)]^2 \Psi[G(x)] \, dG(x)$$

where $F_n(x)$ is the empirical cpf determined by the sample $X_1, \cdots, X_n$; and $X_1', X_2', \cdots, X_n'$ are the ordered sample values. All satisfy (a) and (b) when $\Omega = \Omega' = \Omega_2$. Hence w_n^2, D_n, K_n, and W_n^2 are all statistics in Ω_2 w.r.t. Ω_2.

If for a statistic $S_G(X^{(n)})$ in Ω w.r.t. Ω' there exists a (measurable) function Φ defined on the n-dimensional unit cube and symmetric in its arguments, such

that for any $G \varepsilon \Omega$, $F \varepsilon \Omega'$, we have $S_G(x^{(n)}) \equiv \Phi[G(x_1), \cdots, G(x_n)][\mathcal{P}_F]$, i.e. almost everywhere in the sample space $X^{(n)}$ for the random variable X which has cpf F, then $S_G(X^{(n)})$ is called a *statistic of structure* (d).

If $\Omega = \Omega'$ and $S_G(X^{(n)})$ has the property that $\mathcal{P}_G^{(n)} S_G^{-1}$, the probability distribution of S_G when X has cpf G, is independent of G for all $G \varepsilon \Omega$, then $S_G(X^{(n)})$ is a *distribution-free statistic* in Ω.

If $S_G(X^{(n)})$ is a statistic in $\Omega \subset \Omega^*$ w.r.t. some Ω', then $S_G(X^{(n)})$ is called a *strongly distribution-free statistic* in Ω w.r.t. Ω' if $\mathcal{P}_F^{(n)} S_G^{-1}$ depends only on the function $\tau = FG^{-1}$ for all $G \varepsilon \Omega$ and $F \varepsilon \Omega'$.

In view of the preceding definitions, it can be readily established that

(A) if a statistic in Ω_2 w.r.t. Ω_2 has structure (d) then it is distribution-free in Ω_2;

(B) if a statistic in Ω^* w.r.t. Ω^* is strongly distribution-free, then it is distribution-free in Ω^*; and

(C) if a statistic in Ω^* w.r.t. Ω^* has structure (d), then it is strongly distribution-free.

Further, it is seen that each of the statistics (von Mises, etc.) in the example above is, for properly chosen classes of cpf's, of structure (d); strongly distribution-free and symmetric; and distribution-free. Such also is the case for D_n^+ and D_n^- of Wald and Wolfowitz [21], and Birnbaum, [10]; the spacing statistics of Kimball [17] and Sherman [19]; and most of the other distribution-free statistics in the literature.

Birnbaum and Rubin [12] have shown that there exists a distribution-free statistic which is not strongly distribution-free; but the other two properties always seem to occur together in a statistic. For that reason it is of interest to find the conditions under which the property of having structure (d) is equivalent to being symmetric and strongly distribution-free.

It is known [12] that these two properties are equivalent for statistics in Ω^* w.r.t. Ω^*. In Section 3 it will be shown that the two properties are equivalent for statistics in Ω^* w.r.t. Ω', where Ω' satisfies certain closure and completeness properties.

Before proceeding with the proof of this theorem, it is worthwhile to recall some definitions and results concerning cpf's. This is done below in Section 2.

2. Probability Functions. In view of the nature of the problem, the work will deal primarily with probability spaces on the real line and on the unit interval. For that reason the following classes and sets should be defined.

Let R, $R^{(n)}$, I, $I^{(n)}$, $\mathcal{B}$, $\mathcal{B}^{(n)}$, $\mathcal{B}_I$, and $\mathcal{B}_I^{(n)}$, be respectively, the real line; euclidean n-space; the open unit interval; the n-dimensional open unit cube; and the respective classes of borel subsets of R, $R^{(n)}$, I, $I^{(n)}$.

A cpf, $F(x)$, on R is a non-decreasing, upper semi-continuous function defined on R and such that $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$. A cpf, $H(u)$, on I is a non-decreasing, upper semi-continuous function defined on I and such that $\lim_{u\to1} H(u) = 1$ and $\lim_{u\to0} H(u) = 0$.

It is well known ([2], p. 96) that each cpf on R induces and is induced by a probability distribution on $\mathcal{B}$; similarly each cpf on I induces and is induced by a probability distribution on $\mathcal{B}_I$. Let $\mathcal{P}_F$ denote the probability distribution induced by the cpf $F(x)$; and let $\mathcal{P}_F^{(n)}$ denote *power probability distribution* on the class $\mathcal{B}^{(n)}$ generated by F, i.e. the probability distribution induced by n independent random variables each distributed with cpf F.

If G, $G_1 \varepsilon \Omega^*$, then G, G^{-1}, and G_1G^{-1} are all $1 - 1$ strictly monotone, continuous mappings; and, hence, preserve many of the properties of cpf's and their probability distributions. In fact,

(i) if $F \varepsilon \Omega_0[\Omega_1, \Omega_2, \Omega^*]$ and $G, G_1 \varepsilon \Omega^*$, then

(a) $FG^{-1} \varepsilon \Omega_0(I)[\Omega_1(I), \Omega_2(I), \Omega^*(I)]$ and

(b) $FG^{-1}G_1 \varepsilon \Omega_0[\Omega_1, \Omega_2, \Omega^*]$.

Since the closure property (b) is important in the sequel, it is worthwhile to give the following formal definition.

Ω' is said to be *closed under* Ω if $FG^{-1}G_1 \varepsilon \Omega'$, whenever $F \varepsilon \Omega'$ and G, $G_1 \varepsilon \Omega$. Therefore, one concludes from (i) that Ω_0, Ω_1, Ω_2 and Ω^* are each closed under Ω^*.

Further, it is seen that under such mappings numerical values are preserved in the following sense.

(ii) If $F \varepsilon \Omega_0$ and $G, G_1 \varepsilon \Omega^*$, then $(\alpha)\mathcal{P}_{FG^{-1}}^{(n)}(B) = \mathcal{P}_F^{(n)}G^{-1}(B)$ for all $B \varepsilon \mathcal{B}_I^{(n)}$; and $(\beta)\mathcal{P}_{FG^{-1}G_1}^{(n)}[G_1^{-1}(B)] = \mathcal{P}_{FG^{-1}}^{(n)}(B)$ for all $B \varepsilon \mathcal{B}_I^{(n)}$, where

$$[G(x^{(n)})] = [G(x_1), \cdots, G(x_n)]$$

and $G^{-1}(u, \cdots, u_n) = [G^{-1}(u_1), \cdots, G^{-1}(u_n)]$.

With these preliminary results one can proceed to establish the main theorem.

3. The Main Theorem. As mentioned in the introduction the object here is to demonstrate that for suitable classes of cpf's a statistic is symmetric and distribution-free if and only if it is of structure (d).

If a statistic, S_G, in Ω w.r.t. Ω' is of structure (d), there exists a measurable function Φ defined on $I^{(n)}$ and symmetric in its arguments, such that for any $G \varepsilon \Omega$ and $F \varepsilon \Omega'$, $S_G(x^{(n)}) \equiv \Phi[G(x^{(n)})][\mathcal{P}_F]$.

If A is an arbitrary element of $\mathcal{B}^{(n)}$, then $S_G^{-1}(A) \equiv G^{-1}o\Phi^{-1}(A)$. In view of (ii), then, $\mathcal{P}_FS_G^{-1}(A) \equiv \mathcal{P}_FG^{-1}o\Phi^{-1}(A) \equiv \mathcal{P}_{FG^{-1}}\Phi^{-1}(A)$ providing FG^{-1} is well defined. Clearly, this will be so whenever $G \varepsilon \Omega^*$. Further, S_G is symmetric whenever Φ is. Therefore, one can conclude the following.

LEMMA 1: *If a statistic, S_G, in $\Omega \subset \Omega^*$* w.r.t. Ω' *is of structure* (d), *then S_G is symmetric and strongly distribution-free.*

On the other hand if S_G, a statistic in $\Omega \subset \Omega^*$ w.r.t. Ω', is symmetric and strongly distribution-free, let $\Phi_1 \equiv S_{G_1}\, o\, G_1^{-1}$, where G_1 is an arbitrary fixed element of $\Omega \subset \Omega^*$.

It is clear that Φ_1 is symmetric. Therefore, in order to complete the proof one must demonstrate that $S_G(x^{(n)}) = \Phi_1[G(x^{(n)})][\mathcal{P}_F]$ for all $F \varepsilon \Omega'$ and all

$$G \varepsilon \Omega \subset \Omega^*.$$

Again let A be an arbitrary fixed element of $\mathcal{B}^{(n)}$. Then,

$$\mathcal{P}_F\{\Phi_1[G(x^{(n)})] \varepsilon A\} = \mathcal{P}_F G^{-1} o \Phi_1^{-1}(A) = \mathcal{P}_{FG^{-1}}\Phi_1^{-1}(A) = \mathcal{P}_{FG^{-1}}G_1 o S_{G_1}^{-1}(A)$$
$$= \mathcal{P}_{FG^{-1}G_1}S_{G_1}^{-1}(A) \qquad \text{for all } F \varepsilon \Omega' \text{ and all } G \varepsilon \Omega \subset \Omega^*.$$

Now, if $FG^{-1}G_1 \varepsilon \Omega'$, i.e. if Ω' is closed under $\Omega \subset \Omega^*$, then the fact that S_G is strongly distribution-free guarantees that $\mathcal{P}_{FG^{-1}G_1}S_{G_1}^{-1}(A) = \mathcal{P}_F S_G^{-1}(A)$ since $(FG^{-1}G_1)G_1^{-1} = (F)G^{-1}$. Under these circumstances one sees that

$$\mathcal{P}_F\{\Phi_1[G(x^{(n)})] \varepsilon A\} = \mathcal{P}_F\{S_G(x^{(n)}) \varepsilon A\} \qquad \text{for all } F \varepsilon \Omega',\ G \varepsilon \Omega \subset \Omega^*.$$

These results lead one to the following question. What conditions must the class Ω' satisfy in order that S_G and $\Phi \, o \, G$, which have identical distributions for each $F \varepsilon \Omega'$, be essentially equal? In answering this question, the following definition will be employed.

A class, Ω, of cpf's is said to be *symmetrically complete* if every unbiased, symmetric estimator of zero, with respect to the class of power probability distributions of Ω, is essentially zero, i.e., the conditions (1) f is symmetric; and (2) $\int_{R^{(n)}} f d\, \mathcal{P}_F^{(n)} = 0$ for all $F \varepsilon \Omega$, imply that $f = 0[\mathcal{P}_F^{(n)}]$ for all $F \varepsilon \Omega$.

In terms of this definition, the answer to the question is as follows.

LEMMA 2: *If S and Φ are symmetric measurable functions such that*

$$\mathcal{P}_F^{(n)}\{S \varepsilon A\} = \mathcal{P}_F^{(n)}\{\Phi \varepsilon A\}$$

for all $A \varepsilon \mathcal{B}$ and all $F \varepsilon \Omega'$; and if Ω' is a symmetrically complete class, then

$$S = \Phi[\mathcal{P}_F^{(n)}]$$

for all $F \varepsilon \Omega'$.

PROOF: Let $g(B, x^{(n)})$ be the indicator function of B, i.e.

$$g(B, x^{(n)}) = \begin{cases} 1 \text{ for } x^{(n)} \varepsilon B, \\ 0 \text{ otherwise;} \end{cases}$$

then for each $A \varepsilon \mathcal{B}$ and each $F \varepsilon \Omega'$,

$$\int_{R^{(n)}} [g(S^{-1}(A), x^{(n)}) - g(\Phi^{-1}(A), x^{(n)})]\, d\mathcal{P}_F^{(n)} = \mathcal{P}_F^{(n)}\{S^{-1}(A)\}$$
$$- \mathcal{P}_F^{(n)}\{\Phi^{-1}(A)\} = 0.$$

Since S and Φ are symmetric, $g(S^{-1}(B), x^{(n)})$ and $g(\Phi^{-1}(B), x^{(n)})$ are symmetric, and so is their difference. Because of the completeness property of Ω',

$$g(S^{-1}(B), x^{(n)}) - g(\Phi^{-1}(B), x^{(n)}) = 0$$

and $g(S^{-1}(B), x^{(n)}) = g(\Phi^{-1}(B), x^{(n)})[\mathcal{P}_F^{(n)}]$ for all $F \varepsilon \Omega'$. Consequently,

$$\mathcal{P}_F^{(n)}(S^{-1}(A) \Delta \Phi^{-1}(A)) = 0$$

for all $F \varepsilon \Omega'$ and all $A \varepsilon B$.

[Note: $E \Delta F = (E \cup F) - (E \cap F)$.]

But

$$\mathcal{P}_F^{(n)}(S \neq \Phi) = \mathcal{P}_F^{(n)}(S > \Phi) + \mathcal{P}_F^{(n)}(\Phi > S) \leq$$

$$\sum_{m=-\infty}^{\infty} \sum_{k=1}^{\infty} [\mathcal{P}_F^{(n)}(S > (m/k); \Phi < (m/k)) + \mathcal{P}_F^{(n)}(S < (m/k); \Phi > (m/k))]$$

$$\leq \sum_{m=-\infty}^{\infty} \sum_{k=1}^{\infty} [\mathcal{P}_F^{(n)}((S \leq (m/k))\Delta(\Phi \leq (m/k)))] = 0 \quad \text{for all } F \varepsilon \Omega'.$$

Therefore $S = \Phi[\mathcal{P}_F^{(n)}]$ for all $F \varepsilon \Omega$. The main theorem now follows immediately.

THE MAIN THEOREM. *If S_G is a statistic in Ω* w.r.t. Ω', *then the property of being symmetric and strongly distribution-free is equivalent to having structure* (d), *whenever the following three conditions are fulfilled.*

(α) $\Omega \subset \Omega^*$;

(β) Ω' *is closed under* Ω; *and*

(γ) Ω' *is a symmetrically complete class.*

The next question is: Which classes of statistical interest satisfy the hypotheses of the main theorem?

4. Closed and complete classes. As was previously mentioned one can conclude from (i) that Ω_0, Ω_1, Ω_2 and Ω^* are closed under all subsets of Ω^*. Also, it can be proved that Ω_3, Ω_4, Ω_u and Ω_e do not satisfy that closure property. However, one can verify that Ω_3 is closed under $\Omega_3 \cap \Omega^*$; and that Ω_4 is closed under $\Omega_4 \cap \Omega^*$.

The work of Halmos [16]; Fraser ([14], [15], [1], pp. 23–31); Lehmann ([3], p. 132), and Bell-Blackwell-Breiman [8] establish the fact that Ω_0, Ω_1, Ω_2, Ω_3, Ω_4, Ω_u and Ω_e are symmetrically complete. (It should be mentioned here that a class of cpf's is symmetrically complete if and only if the order statistic is a complete statistic with respect to the class of power probability distributions of the given class of cpf's.)

Therefore, Ω_0, Ω_1, Ω_2, Ω_3 and Ω_4 satisfy both the completeness and closure hypotheses of the main theorem. Consequently, the following corollary to the main theorem is valid.

COROLLARY: *If S_G is a statistic in Ω* w.r.t. Ω', *then the property of being symmetric and strongly distribution-free is equivalent to having structure* (d) *for each of the following cases.*

(1) $\Omega \subset \Omega^*$ and $\Omega' = \Omega_0$;

(2) $\Omega \subset \Omega^*$ and $\Omega' = \Omega_1$;

(3) $\Omega \subset \Omega^*$ and $\Omega' = \Omega_2$;

(4) $\Omega \subset \Omega^*$ and $\Omega' = \Omega^*$;

(5) $\Omega = \Omega_3 \cap \Omega^*$ and $\Omega' = \Omega_3$; and

(6) $\Omega = \Omega_4 \cap \Omega^*$ and $\Omega' = \Omega_4$.

REFERENCES

[1] D. A. S. FRASER, *Non-parametric Methods in Statistics*, New York, John Wiley and Sons, 1957.

[2] M. Loève, *Probability Theory*, New York, D. van Nostrand, 1955.

[3] E. Lehmann, *Testing Statistical Hypotheses*, New York, John Wiley and Sons, 1959.

[4] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 183–212.

[5] C. B. Bell, "Application of distribution-free statistics to some problems in missile design and production," Douglas (Aircraft Co.), Santa Monica Report SM 18396 (1954).

[6] C. B. Bell, "On the structure of algebras and homomorphisms," *Proc. Amer. Math. Soc.*, Vol. 7 (1956), pp. 483–492.

[7] C. B. Bell, "On the structure of stochastic independence," *Ill. J. Math.*, Vol. 2 (1958), pp. 415–424.

[8] C. B. Bell, D. Blackwell and L. Breiman, "A note on the completeness of order statistics," *Ann. Math. Stat.*, Vol. 31 (1960), pp. 794–797.

[9] Z. W. Birnbaum, "Numerical tabulation of the distribution of Kolmogoroff's statistic for finite sample size," *J. Amer. Stat. Assoc.*, Vol. 47 (1952), pp. 425–441.

[10] Z. W. Birnbaum and F. H. Tingey, "One-sided confidence contours for probability distribution functions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 592–596.

[11] Z. W. Birnbaum, "Distribution-free tests for continuous distribution functions," *Ann. Math. Stat*, Vol. 24 (1953), pp. 1–8.

[12] Z. W. Birnbaum and H. Rubin, "On distribution-free statistics," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 593–598.

[13] W. Feller, "On the Kolmogoroff-Smirnoff limit theorems for empirical distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 177–189.

[14] D. A. S. Fraser, "Completeness of order statistics," *Can. J. Math.*, Vol. 6 (1953), pp. 42–45.

[15] D. A. S. Fraser, "Non-parametric theory: scale and location parameters," *Can. J. Math.*, Vol. 6 (1953), pp. 46–68.

[16] P. R. Halmos, "The theory of unbiased estimation," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 34–43.

[17] B. F. Kimball, "Some basic theorems for developing tests of fit for the case of the non-parametric probability distribution function, I," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 540–548.

[18] H. Scheffé, "On a measure problem arising in the theory of non-parametric tests," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 227–233.

[19] B. Sherman, "A random variable related to the spacing of sample values," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 339–361.

[20] N. Smirnov, "Table for estimating the goodness of fit of empirical distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 279–281.

[21] A. Wald and J. Wolfowitz, "Confidence limits for continuous distribution functions," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 105–118.

SMALL SAMPLE DISTRIBUTIONS FOR MULTI-SAMPLE STATISTICS OF THE SMIRNOV TYPE

By Z. W. Birnbaum and R. A. Hall[1]

University of Washington

1. Introduction and Summary. Let

$$X_1^{(i)}, X_2^{(i)}, \cdots, X_{n_i}^{(i)}, \qquad i = 1, 2, \cdots, c, \tag{1.1}$$

be samples of c independent random variables $X^{(i)}$ with continuous cumulative distribution functions $F^{(i)}$, and let

$$\begin{aligned} F^{*(i)}(x) &= 0 && x < X_1^{(i)} \\ F^{*(i)}(x) &= k/n_i && X_k^{(i)} \leqq x < X_{k+1}^{(i)},\ 1 \leqq k < n_i \\ F^{*(i)}(x) &= 1 && X_{n_i}^{(i)} \leqq x \end{aligned} \tag{1.2}$$

be the corresponding c empirical distribution functions. We define the statistics

$$D(n_1, n_2, \cdots, n_c) = \sup_{\substack{x,i,j \\ (i,j=1,2,\cdots,c)}} | F^{*(i)}(x) - F^{*(j)}(x) | \tag{1.3}$$

and

$$D^+(n_1, n_2, \cdots, n_c) = \sup_{\substack{x,i,j \\ (i<j;\, i,j=1,2,\cdots,c)}} [F^{*(i)}(x) - F^{*(j)}(x)]. \tag{1.4}$$

The well known Kolmogorov-Smirnov statistics $D(m, n)$ and $D^+(m, n)$ are special cases of (1.3) and (1.4), respectively, with $c = 2$, $n_1 = m$, $n_2 = n$.

The exact small sample distribution, under the null hypothesis

$$F^{(i)} = F^{(j)} \qquad \text{for all } i, j = 1, 2, \cdots, c, \tag{1.5}$$

of the statistics defined by (1.3) and (1.4) for any number c of samples, and for any sample sizes $n_1, n_2, \cdots, n_c$, can be obtained by solving simple difference equations which lend themselves to programming for machine computation. Using this procedure, tables of values of

$$P[D(n, n, n) \leqq r], \qquad P[D(n, n) \leqq r], \qquad P[D^+(n, n) \leqq r]$$

were computed for selected values of n between 1 and 40 and of $r = k/n$, $k = 1, 2, \cdots, n$.

Furthermore, the inequalities

$$P[D(n, n, \cdots, n) \leqq r] \geqq 1 - [c(c-1)/2]P[D(n, n) > r]$$

$$P[D(n, n, \cdots, n) \leqq r] \geqq 1 - [c(c-1)(c-2)/6]P[D(n, n, n) > r]$$

Received August 15, 1959.

[1] This research was sponsored by the Office of Naval Research.

are noted, which may be useful for values of $c \geq 4$ for which tables are not available.

2. The general difference equations. The set of all possible values of the c-dimensional random variable $[F^{*(1)}(x), F^{*(2)}(x), \cdots, F^{*(c)}(x)]$, for fixed x, consists of the points

$$(2.1) \qquad (k_1/n_1, k_2/n_2, \cdots k_c/n_c) \qquad k_i = 1, 2, \cdots n_i; \quad i = 1, 2, \cdots, c,$$

of the c dimensional unit cube.

By the transformation $y_i = n_i x_i$ the c dimensional unit cube is transformed into the c-dimensional rectangular prism with sides $n_1, n_2, \cdots n_c$, and the points (2.1) are transformed into the points

$$(2.2) \qquad (k_1, k_2, \cdots, k_c) \qquad k_i = 1, 2, \cdots n_i; \quad i = 1, 2, \cdots, c.$$

Under the null hypothesis (1.5) the c samples may be considered as c successive drawings of $n_1, n_2, \cdots n_c$ observations from the same population, with equal probabilities of each of the $N!$ ways of drawing the ordered sample of size N, where

$$(2.4) \qquad N = n_1 + n_2 + \cdots + n_c.$$

The points (2.2) may be interpreted as being obtained in the following manner: the sample values $X_k^{(i)}$, $k = 1, 2, \cdots, n_i$; $i = 1, 2, \cdots, c$, are observed and plotted on the x-axis. It is agreed that, as one moves along the x-axis from $-\infty$ to $+\infty$, the coordinate k_i of the point (2.2) is increased by a unit whenever a value $X_k^{(i)}$ is crossed. By this procedure one obtains a path through points of the form (2.2), starting at $(0, 0, \cdots, 0)$ and ending at $(n_1, n_2, \cdots, n_c)$, and each set of the c samples determines such a path. Under the null hypothesis (1.5) all these paths are equally probable, and their number is clearly

$$(2.5) \qquad Q(n_1, n_2, \cdots, n_c) = N!/(n_1!n_2! \cdots n_c!).$$

We define, generally,

$$(2.6) \qquad \begin{aligned} &Q(k_1, k_2, \cdots, k_c) \\ &\qquad = \text{number of paths from } (0, 0, \cdots, 0) \text{ to } (k_1, k_2, \cdots, k_c) \end{aligned}$$

for any non-negative integers $k_1, k_2, \cdots, k_c$. The function Q satisfies the difference equation

$$(2.7) \qquad \begin{aligned} Q(k_1, k_2, \cdots, k_c) &= Q(k_1 - 1, k_2, \cdots, k_c) \\ &\quad + Q(k_1, k_2 - 1, \cdots, k_c) \\ &\quad \cdots \\ &\quad + Q(k_1, k_2, \cdots, k_c - 1) \end{aligned}$$

since the number of ways of getting from $(0, 0, \cdots, 0)$ to $(k_1, k_2, \cdots, k_c)$ is evidently the sum of the numbers of ways of getting to points from which

$(k_1, k_2, \cdots, k_c)$ can be reached in one step. We have for Q the initial condition

$$Q(0, 0, \cdots, 0) = 1. \tag{2.8}$$

To compute all values of Q for $0 \leqq k_i \leqq n_i$, $i = 1, 2, \cdots, c$, one may start with (2.8) and use (2.7) recursively, a procedure which can be programmed for an electronic computer.

Let now R be a given set of points of the form (2.2), and let

$$\begin{aligned} Q(k_1, k_2, \cdots, k_c; R) = {} & \text{number of paths from } (0, 0, \cdots 0) \text{ to} \\ & (k_1, k_2, \cdots, k_c) \text{ which do not pass through any points in } R. \end{aligned} \tag{2.9}$$

Again the difference equation

$$\begin{aligned} Q(k_1, k_2, \cdots, k_c; R) = {} & Q(k_1 - 1, k_2, \cdots, k_c; R) \\ & + Q(k_1, k_2 - 1, \cdots, k_c; R) \\ & \cdots \\ & + Q(k_1, k_2, \cdots, k_c - 1; R) \end{aligned} \tag{2.10}$$

is satisfied, and can be solved recursively under condition (2.8) and the additional conditions

$$Q(k_1, k_2, \cdots, k_c; R) = 0 \quad \text{for} \quad (k_1, k_2, \cdots, k_c) \text{ in } R. \tag{2.10.1}$$

This, again, is an algorithm which can be programmed for an electronic computer but the program must now, among others, contain the instruction for the computer to decide at each point $(k_1, k_2, \cdots, k_c)$ whether it belongs to R or not.

We now define

$$P_R(n_1, n_2, \cdots, n_c) = Q(n_1, n_2, \cdots, n_c; R)/Q(n_1, n_2, \cdots, n_c), \tag{2.11}$$

the probability that, under the null hypothesis (1.5), the samples determine a path from $(0, 0, \cdots, 0)$ to $(n_1, n_2, \cdots, n_c)$ which does not pass through any point of R.

If, for a given set R, we agree to reject the hypothesis (1.5) whenever the samples determine a path containing points in R, then $1 - P_R$ is the probability of an error of the first kind, i.e. of rejecting the hypothesis when it is true. The tabulation of P_R is manageable for reasonable numbers of samples c and sample sizes $n_1, n_2, \cdots, n_c$, and for R such that one can program for the computer a rule for deciding whether a point is in R or not.

For the Kolmogorov-Smirnov statistic $D(n_1, n_2)$ the sets R are usually defined by $D(n_1, n_2) > r$, which is equivalent with

$$R_r: \quad |n_2k_1 - n_1k_2| > n_1n_2r, \tag{2.12}$$

and for the one-sided statistic $D^+(n_1, n_2)$ by $D^+(n_1, n_2) > r$, equivalent with

$$R_r': \quad n_2k_1 - n_1k_2 > n_1n_2r. \tag{2.12.1}$$

For $n_1 = n_2$, (2.12) and (2.12.1) become $|k_1 - k_2| > nr$ and $k_1 - k_2 > nr$, respectively.

Analogous multi-sample tests can be defined by using the statistics (1.3) or (1.4) and the regions of rejection

$$D(n_1, n_2, \cdots, n_c) > r \quad \text{and} \quad D^+(n_1, n_2, \cdots, n_c) > r,$$

respectively. The corresponding sets R are

$$\sup_{(i,j=1,\cdots,c)} |n_j k_i - n_i k_j| > r, \tag{2.13}$$

and

$$\sup_{(i<j)} (n_j k_i - n_i k_j) > r \tag{2.13.1}$$

respectively.

It may be noted that the computations involved in tabulating

$$P_R(n_1, n_2, \cdots, n_c)$$

would not be much more difficult to program and more time-consuming if (2.13) or (2.13.1) were replaced by more general sets R such as

$$|n_j k_i - n_i k_j| > f(k_1, k_2, \cdots, k_c)$$

for some reasonably simple function f.

The tables described in the next section were computed by using difference equations (2.7) and (2.10). It should be stated that these difference equations have been well known and used for the case $c = 2$, and that closed expressions for $P_R(n_1, n_2)$ were obtained in special cases, e.g. by Gnedenko and Korolyuk [3] and by Drion [2]. An excellent summary of the history of these methods may be found in the paper by Hodges [4]. A more recent paper by David [1] contains the derivation of the small-sample distribution and the asymptotic distribution of the statistic

$$\operatorname{Max}\{\sup_{(x)} [F^{*(2)}(x) - F^{*(1)}(x)], \quad \sup_{(x)} [F^{*(3)}(x) - F^{*(2)}(x)],$$
$$\sup_{(x)} [F^{*(1)}(x) - F^{*(3)}(x)]\}.$$

3. Tables. Table 1 contains the probabilities $P[D(n, n, n) \leqq r]$ for $n = 1$ (1) 20 (2) 40 and consecutive integer values nr such that the probabilities for each n range from less than .90 to more than .995.

Table 2 contains the probabilities $P[D(n, n) \leqq r)$ for $n = 1$ (1) 40 and $nr = 1$ (1) min $(n, 20)$.

Table 3 contains the probabilities $P[D^+(n, n) \leqq r]$ for $n = 1$ (1) 40 and $nr = 1$ (1) min $(n, 20)$.

All probabilities are given to six decimal places. Conservative error estimates assure an error $<5.10^{-6}$ throughout Table 1 and error $<(2.3)10^{-6}$ throughout Tables 2 and 3, but the actual errors are likely to be much smaller.

TABLE 1

$P[D(n, n, n) \leq r]$

r \ n	1	2	3	4	5	6	7	8	9	10
1	1 000000	0 400000	0 128571	0 037402	0 010275	0 002719	0 000701	0 000177	0 000044	0 000010
2		1 000000	0 771428	0 539220	0 355929	0 226374	0 140271	0 085256	0 051053	0 030213
3			1 000000	0 926406	0 811093	0 684084	0 562086	0 453012	0 359715	0 282279
4				1 000000	0 978188	0 932164	0 868227	0 793917	0 715417	0 637148
5					1 000000	0 993829	0 977501	0 950288	0 913501	0 869301
6						1 000000	0 998303	0 992915	0 982475	0 966446
7							1 000000	0 999541	0 997847	0 994114
8								1 000000	0 999877	0 999362
9									1 000000	0 999967
10										1 000000
	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	0 000002	0 000000	0 000000							
2	0 017709	0 010297	0 005948							
3	0 219397	0 169169	0 129569							
4	0 562027	0 491832	0 427525							
5	0 819975	0 767590	0 713862							
6	0 944960	0 918575	0 888073	0 854312	0 818130	0 780302	0 741507	0 702328	0 663250	0 624670
7	0 987711	0 978261	0 965629	0 949882	0 931228	0 909969	0 886458	0 861064	0 834155	0 806081
8	0 998093	0 995691	0 991826	0 986249	0 978802	0 969415	0 958096	0 944916	0 929988	0 913459
9	0 999815	0 999399	0 998539	0 997044	0 994729	0 991436	0 987039	0 981452	0 974624	0 966539
10	0 999991	0 999947	0 999814	0 999518	0 998964	0 998049	0 996668	0 994723	0 992126	0 988805
11	1 000000	0 999997	0 999985	0 999943	0 999844	0 999646	0 999298	0 998744	0 997922	0 996773
12		1 000000	0 999999	0 999995	0 999983	0 999950	0 999881	0 999754	0 999539	0 999205
13			1 000000	0 999999	0 999998	0 999994	0 999984	0 999961	0 999915	0 999834
14				1 000000	0 999999	0 999999	0 999998	0 999995	0 999987	0 999971
15					1 000000	0 999999	0 999999	0 999999	0 999998	0 999996
	22	**24**	**26**	**28**	**30**	**32**	**34**	**36**	**38**	**40**
8	0 876276	0 834896	0 790312	0 744128	0 697257					
9	0 946679	0 922382	0 893835	0 862177	0 828007	0 792099	0 754883	0 717132	0 679257	0 641658
10	0 979784	0 967557	0 951678	0 932761	0 910963	0 886657	0 860253	0 832162	0 802779	0 772473
11	0 993268	0 987984	0 980202	0 970227	0 957886	0 943250	0 926465	0 907727	0 887261	0 865309
12	0 998039	0 996114	0 992702	0 988029	0 981779	0 973852	0 964215	0 952885	0 939929	0 925445
13	0 999504	0 998965	0 997584	0 995632	0 992789	0 988907	0 983879	0 977629	0 970120	0 961345
14	0 999892	0 999844	0 999284	0 998556	0 997392	0 995668	0 993276	0 990117	0 986113	0 981208
15	0 999980	0 999928	0 999811	0 999569	0 999139	0 998444	0 997404	0 995937	0 993968	0 991430
16					0 999741	0 999486	0 999073	0 998447	0 997552	0 996333

TABLE 2
$P[D(n, n) \leq r]$

nr	n: 1	2	3	4	5	6	7	8	9	10
1	1 000000	0 666666	0 400000	0 228571	0 126984	0 069264	0 037296	0 019891	0 010530	0 005542
2		1 000000	0 900000	0 771428	0 642857	0 525974	0 424825	0 339860	0 269888	0 213070
3			1 000000	0 971428	0 920634	0 857142	0 787878	0 717327	0 648292	0 582476
4				1 000000	0 992063	0 974025	0 946969	0 912975	0 874125	0 832178
5					1 000000	0 997835	0 991841	0 981351	0 966433	0 947552
6						1 000000	0 999417	0 997513	0 993706	0 987659
7							1 000000	0 999844	0 999259	0 997943
8								1 000000	0 999958	0 999783
8									1 000000	0 999989
10										1 000000
	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	0 002903	0 001514	0 000787	0 000408	0 000211	0 000109	0 000056	0 000028	0 000014	0 000007
2	0 167412	0 131018	0 102194	0 079484	0 061668	0 047743	0 036892	0 028460	0 021922	0 016863
3	0 520849	0 463902	0 411803	0 364515	0 321861	0 283588	0 249392	0 218952	0 191938	0 168030
4	0 788523	0 744224	0 700079	0 656679	0 614453	0 573706	0 534647	0 497409	0 462071	0 428664
5	0 925339	0 900453	0 873512	0 845065	0 815583	0 785465	0 755040	0 724581	0 694310	0 664409
6	0 979260	0 968563	0 955727	0 940970	0 924535	0 906673	0 887622	0 867606	0 846826	0 825466
7	0 995633	0 992140	0 987350	0 981217	0 973751	0 965002	0 955047	0 943981	0 931910	0 918942
8	0 999345	0 998503	0 997125	0 995100	0 992344	0 988800	0 984439	0 979252	0 973250	0 966458
9	0 999937	0 999795	0 999500	0 998979	0 998162	0 996984	0 995389	0 993331	0 990776	0 987701
10	0 999997	0 999982	0 999937	0 999836	0 999646	0 999329	0 998847	0 998160	0 997232	0 996032
11	1 000000	0 999999	0 999995	0 999981	0 999947	0 999880	0 999761	0 999570	0 999285	0 998884
12		1 000000	0 999999	0 999998	0 999994	0 999983	0 999960	0 999916	0 999843	0 999729
13			1 000000	1 000000	0 999999	0 999998	0 999994	0 999987	0 999971	0 999944
14				1 000000	1 000000	0 999999	0 999999	0 999998	0 999995	0 999990
15					1 000000	1 000000	1 000000	0 999999	0 999999	0 999998
16						1 000000	1 000000	1 000000	0 999999	0 999999
17							1 000000	1 000000	1 000000	1 000000
18								1 000000	1 000000	1 000000
19									1 000000	1 000000
20										1 000000
	21	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
1	0 000003	0 000001	0 000001	0 000000	0 000000	0 000000	0 000000	0 000000	0 000000	0 000000
2	0 012955	0 009942	0 007622	0 005838	0 004468	0 003417	0 002611	0 001993	0 001521	0 001160
3	0 146921	0 128321	0 111963	0 097599	0 085006	0 073980	0 064337	0 055914	0 048563	0 042153
4	0 397187	0 367613	0 339809	0 313982	0 289796	0 267262	0 246302	0 226833	0 208772	0 192036
5	0 635020	0 606260	0 578218	0 550963	0 524546	0 499004	0 474362	0 450633	0 427822	0 405029

TABLE 2—(*Continued*)

m	n									
	21	22	23	24	25	26	27	28	29	30
6	0 803687	0 781631	0 759421	0 737166	0 714957	0 692876	0 670992	0 649361	0 628035	0 607054
7	0 905183	0 890738	0 875705	0 860177	0 844239	0 827971	0 811443	0 794721	0 777865	0 760926
8	0 958911	0 950653	0 941731	0 932196	0 922101	0 911498	0 900437	0 888969	0 877140	0 864996
9	0 984094	0 979952	0 975279	0 970086	0 964388	0 958206	0 951561	0 944480	0 936988	0 929112
10	0 994532	0 992710	0 990548	0 988034	0 985162	0 981927	0 978330	0 974375	0 970069	0 965419
11	0 998343	0 997641	0 996759	0 995679	0 994385	0 992865	0 991109	0 989109	0 986859	0 984356
12	0 999561	0 999326	0 999009	0 998598	0 998079	0 997439	0 996666	0 995750	0 994681	0 993451
13	0 999899	0 999831	0 999732	0 999594	0 999409	0 999167	0 998861	0 998482	0 998020	0 997469
14	0 999980	0 999963	0 999936	0 999895	0 999837	0 999756	0 999647	0 999505	0 999325	0 999100
15	0 999996	0 999993	0 999986	0 999976	0 999960	0 999936	0 999901	0 999853	0 999790	0 999706
16	0 999999	0 999998	0 999997	0 999995	0 999991	0 999985	0 999975	0 999961	0 999940	0 999912
17	1 000000	0 999999	0 999999	0 999999	0 999998	0 999996	0 999994	0 999990	0 999984	0 999976
18	1 000000	1 000000	1 000000	0 999999	0 999999	0 999999	0 999998	0 999998	0 999996	0 999994
19	1 000000	1 000000	1 000000	1 000000	0 999999	0 999999	0 999999	0 999999	0 999999	0 999998
20	1 000000	1 000000	1 000000	1 000000	1 000000	1 000000	0 999999	0 999999	0 999999	0 999999
	31	32	33	34	35	36	37	38	39	40
1	0 000000	0 000000	0 000000	0 000000	0 000000	0 000000	0 000000	0 000000	0 000000	0 000000
2	0 000884	0 000674	0 000513	0 000390	0 000297	0 000226	0 000171	0 000130	0 000099	0 000075
3	0 036570	0 031710	0 027482	0 023808	0 020615	0 017844	0 015440	0 013354	0 011546	0 009980
4	0 176546	0 162222	0 148989	0 136773	0 125505	0 115119	0 105553	0 096746	0 088644	0 081194
5	0 384946	0 364860	0 345656	0 327315	0 309815	0 293133	0 277243	0 262120	0 247737	0 234068
6	0 586454	0 566263	0 546505	0 527197	0 508355	0 489989	0 472106	0 454713	0 437810	0 421399
7	0 743954	0 726991	0 710076	0 693241	0 676518	0 659934	0 643511	0 627272	0 611234	0 595412
8	0 852579	0 839930	0 827085	0 814080	0 800946	0 787713	0 774409	0 761059	0 747686	0 734312
9	0 920879	0 912317	0 903453	0 894313	0 884922	0 875305	0 865485	0 855485	0 845325	0 835027
10	0 960438	0 955137	0 949530	0 943629	0 937451	0 931011	0 924322	0 917402	0 910264	0 902925
11	0 981599	0 978588	0 975325	0 971814	0 968060	0 964067	0 959843	0 955395	0 950731	0 945858
12	0 992054	0 990483	0 988735	0 986806	0 984695	0 982400	0 979921	0 977260	0 974418	0 971396
13	0 996821	0 996069	0 995206	0 994228	0 993128	0 991904	0 990551	0 989067	0 987450	0 985698
14	0 998825	0 998494	0 998102	0 997644	0 997113	0 996507	0 995820	0 995049	0 994189	0 993239
15	0 999600	0 999466	0 999302	0 999104	0 998868	0 998589	0 998265	0 997891	0 997464	0 996981
16	0 999875	0 999825	0 999762	0 999683	0 999586	0 999467	0 999325	0 999156	0 998958	0 998729
17	0 999964	0 999947	0 999925	0 999896	0 999859	0 999812	0 999754	0 999683	0 999598	0 999496
18	0 999990	0 999985	0 999978	0 999968	0 999955	0 999938	0 999916	0 999888	0 999854	0 999812
19	0 999997	0 999996	0 999994	0 999991	0 999987	0 999981	0 999973	0 999963	0 999950	0 999934
20	0 999999	0 999999	0 999998	0 999997	0 999996	0 999994	0 999992	0 999988	0 999984	0 999978

TABLE 3

$P[D^+(n, n) \leq r]$

nr	n = 1	2	3	4	5	6	7	8	9	10
0	500000	333333	250000	200000	166666	142857	125000	111111	100000	90909
1	1 000000	0 833333	0 700000	0 600000	0 523809	0 464285	0 416666	0 377777	0 345454	0 318181
2		1 000000	0 950000	0 885714	0 821428	0 761904	0 708333	0 660606	0 618181	0 580419
3			1 000000	0 985714	0 960317	0 928571	0 893939	0 858585	0 823776	0 790209
4				1 000000	0 996031	0 987012	0 973484	0 956487	0 937062	0 916083
5					1 000000	0 998917	0 995920	0 990675	0 983216	0 973776
6						1 000000	0 999708	0 998756	0 996853	0 993829
7							1 000000	0 999922	0 999629	0 998971
8								1 000000	0 999979	0 999891
9									1 000000	0 999994
10										1 000000

nr	11	12	13	14	15	16	17	18	19	20
0	83333	76923	71428	66666	62500	58823	55555	52631	50000	47619
1	0 294871	0 274725	0 257142	0 241666	0 227941	0 215686	0 204678	0 194736	0 185714	0 177489
2	0 546703	0 516483	0 489285	0 464705	0 442401	0 422084	0 403508	0 386466	0 370779	0 356295
3	0 758241	0 728021	0 699579	0 672875	0 647832	0 624354	0 602339	0 581681	0 562281	0 544042
4	0 894230	0 872010	0 849789	0 827829	0 806308	0 785345	0 765018	0 745371	0 726425	0 708187
5	0 962669	0 950226	0 936753	0 922523	0 907765	0 892672	0 877401	0 862076	0 846798	0 831646
6	0 989630	0 984281	0 977863	0 970485	0 962267	0 953336	0 943808	0 933796	0 923399	0 912705
7	0 997816	0 996070	0 993675	0 990608	0 986875	0 982500	0 977523	0 971990	0 965955	0 959470
8	0 999672	0 999251	0 998562	0 997550	0 996172	0 994400	0 992219	0 989626	0 986625	0 983229
9	0 999968	0 999897	0 999750	0 999489	0 999081	0 998492	0 997694	0 996665	0 995388	0 993850
10	0 999998	0 999991	0 999968	0 999918	0 999823	0 999664	0 999423	0 999080	0 998616	0 998016
11	1 000000	0 999999	0 999997	0 999990	0 999973	0 999940	0 999880	0 999785	0 999642	0 999442
12		1 000000	0 999999	0 999999	0 999997	0 999991	0 999980	0 999958	0 999921	0 999864
13			1 000000	1 000000	0 999999	0 999999	0 999997	0 999993	0 999985	0 999972
14				1 000000	1 000000	1 000000	0 999999	0 999999	0 999997	0 999995
15					1 000000	1 000000	1 000000	0 999999	0 999999	0 999999
16						1 000000	1 000000	1 000000	1 000000	0 999999
17							1 000000	1 000000	1 000000	1 000000
18								1 000000	1 000000	1 000000
19									1 000000	1 000000
20										1 000000

TABLE 3—(*Continued*)

m'	n = 21	22	23	24	25	26	27	28	29	30
0	45454	43478	41666	40000	38461	37037	35714	34482	33333	32258
1	0 169960	0 163043	0 156666	0 150769	0 145299	0 140211	0 135467	0 131034	0 126881	0 122983
2	0 342885	0 330434	0 318846	0 308034	0 297924	0 288451	0 279556	0 271190	0 263306	0 255865
3	0 526877	0 510702	0 495441	0 481025	0 467390	0 454479	0 442237	0 430617	0 419574	0 409069
4	0 690650	0 673801	0 657621	0 642086	0 627173	0 612856	0 599108	0 585903	0 573216	0 561022
5	0 816681	0 801950	0 787488	0 773321	0 759466	0 745936	0 732738	0 719875	0 707348	0 695154
6	0 901793	0 890731	0 879577	0 868380	0 857183	0 846022	0 834926	0 823922	0 813028	0 802262
7	0 952590	0 945365	0 937846	0 930077	0 922100	0 913953	0 905672	0 897287	0 888826	0 880316
8	0 979455	0 975326	0 970865	0 966097	0 961050	0 955747	0 950216	0 944479	0 938562	0 932486
9	0 992047	0 989976	0 987639	0 985043	0 982194	0 979103	0 975780	0 972239	0 968493	0 964555
10	0 997266	0 996355	0 995274	0 994017	0 992581	0 990963	0 989165	0 987187	0 985034	0 982709
11	0 999171	0 998820	0 998379	0 997839	0 997192	0 996432	0 995554	0 994554	0 993429	0 992178
12	0 999780	0 999663	0 999504	0 999299	0 999039	0 998719	0 998333	0 997875	0 997340	0 996725
13	0 999949	0 999915	0 999866	0 999797	0 999704	0 999583	0 999430	0 999241	0 999010	0 998734
14	0 999990	0 999981	0 999968	0 999948	0 999918	0 999878	0 999823	0 999752	0 999662	0 999550
15	0 999998	0 999996	0 999993	0 999988	0 999980	0 999968	0 999950	0 999926	0 999895	0 999853
16	0 999999	0 999999	0 999998	0 999997	0 999995	0 999992	0 999987	0 999980	0 999970	0 999956
17	1 000000	1 000000	0 999999	0 999999	0 999999	0 999998	0 999997	0 999995	0 999992	0 999988
18	1 000000	1 000000	1 000000	1 000000	0 999999	0 999999	0 999999	0 999998	0 999998	0 999997
19	1 000000	1 000000	1 000000	1 000000	1 000000	1 000000	0 999999	0 999999	0 999999	0 999999
20	1 000000	1 000000	1 000000	1 000000	1 000000	1 000000	0 999999	0 999999	0 999999	0 999999

m'	31	32	33	34	35	36	37	38	39	40
0	31250	30303	29411	28571	27777	27027	26315	25641	24999	24390
1	0 119318	0 115864	0 112605	0 109523	0 106606	0 103840	0 101214	0 098717	0 096341	0 094076
2	0 248830	0 242169	0 235854	0 229858	0 224158	0 218732	0 213562	0 208630	0 203919	0 199416
3	0 399064	0 389525	0 380422	0 371726	0 363411	0 355454	0 347832	0 340525	0 333514	0 326782
4	0 549298	0 538019	0 527164	0 516712	0 506644	0 496940	0 487582	0 478554	0 469840	0 461425
5	0 683290	0 671750	0 660528	0 649616	0 639007	0 628693	0 618666	0 608916	0 599435	0 590215
6	0 791638	0 781167	0 770856	0 760713	0 750743	0 740949	0 731333	0 721895	0 712638	0 703559
7	0 871777	0 863229	0 854689	0 846173	0 837693	0 829262	0 820888	0 812582	0 804349	0 796197
8	0 926272	0 919939	0 913505	0 906988	0 900402	0 893763	0 887081	0 880371	0 873642	0 866904
9	0 960438	0 956157	0 951724	0 947152	0 942454	0 937643	0 932729	0 927724	0 922638	0 917480
10	0 980219	0 977568	0 974764	0 971814	0 968725	0 965504	0 962160	0 958699	0 955130	0 951459
11	0 990799	0 989294	0 987662	0 985907	0 984029	0 982033	0 979921	0 977697	0 975365	0 972929
12	0 996027	0 995241	0 994367	0 993403	0 992347	0 991200	0 989960	0 988630	0 987200	0 985698
13	0 998410	0 998034	0 997603	0 997113	0 996564	0 995952	0 995275	0 994533	0 993725	0 992849
14	0 999412	0 999247	0 999051	0 998821	0 998556	0 998253	0 997910	0 997524	0 997094	0 996619
15	0 999800	0 999733	0 999651	0 999552	0 999434	0 999294	0 999132	0 998945	0 998732	0 998490
16	0 999937	0 999912	0 999881	0 999841	0 999793	0 999733	0 999662	0 999578	0 999479	0 999364
17	0 999982	0 999973	0 999962	0 999948	0 999929	0 999906	0 999877	0 999841	0 999799	0 999748
18	0 999995	0 999992	0 999989	0 999984	0 999977	0 999969	0 999958	0 999944	0 999927	0 999905
19	0 999998	0 999998	0 999997	0 999995	0 999993	0 999990	0 999986	0 999981	0 999975	0 999967
20	0 999999	0 999999	0 999999	0 999998	0 999998	0 999997	0 999996	0 999994	0 999992	0 999989

Table 2 is an extension of the table given by Massey [5]. Tables 1 and 3 appear to be new. Table 3 could also have been computed by a method due to Drion [2].

The computations were programmed for and carried out on the IBM 650 of the Research Computer Laboratory of the University of Washington. The authors wish to express their sincere appreciation to Professor D. B. Dekker for his generous help in planning and performing these computations.

4. Case of $c > 3$. With increasing number of samples c, the computations are not more complicated in structure but quickly become prohibitive in view of the increasing demand on the storage capacity of the computer and the number of additions required. Tabulations similar to those presented in the preceding section, while feasible, would hardly be worth the effort for many values of $c > 3$. Should exact tests based on the statistics $D(n_1, n_2, \cdots, n_c)$, $D^+(n_1, n_2, \cdots, n_c)$ be practically needed then, instead of computing tables, it may be preferable to prepare a program for an electronic computer which, for given sample values, would calculate the single probability needed in every specific case.

Lacking such a program, one may for $c \geqq 3$ make use of the following simple inequalities.

One clearly has, for $c \geqq 3$,

$$\begin{aligned} P[D(n_1, n_2, \cdots, n_c) \leqq r] &= P[\max_{1 \leqq i < j \leqq c} \sup_x | F^{*(i)}(x) - F^{*(j)}(x) | \leqq r] \\ &= 1 - P[\sup_x | F^{*(i)}(x) - F^{*(j)}(x) | > r \text{ for some } i < j] \\ &\geqq 1 - \sum\sum_{1 \leqq i < j \leqq c} P[D(n_i, n_j) > r] \end{aligned}$$

and, for $n_1 = n_2 = \cdots = n_c$, $c \geqq 3$,

$$P[D(n, n, \cdots, n) \leqq r] \geqq 1 - [c(c-1)/2]P[D(n, n) > r]. \tag{4.1}$$

For $c \geqq 4$, one similarly obtains

$$P[D(n, n, \cdots, n) \leqq r] \geqq 1 - [c(c-1)(c-2)/6]P[D(n, n, n) > r]. \tag{4.2}$$

These inequalities make it possible to use the statistic $D(n, n, \cdots, n)$ for testing the hypothesis (1.5) using only Table 1 or Table 2, whichever yields a greater value for the right side of (4.2) or (4.1), respectively. The test will be conservative, i.e. the probability of error of the first kind is less than that obtained from (4.2) or (4.1), but for the conventional "significance levels" and c not too large the right sides in both inequalities should be close approximations to the left side.

Similar inequalities are easily obtained for the statistic $D^+(n, n, \cdots, n)$.

It has been pointed out quite strikingly by Hodges [4] that, for $c = 2$, asymptotic expressions such as that due to Smirnov [6, 7] are inaccurate even for fairly large values of n, to an extent which makes it inadvisable to use them. It appears, therefore, rather doubtful that good approximations can be found for $c > 3$, and as long as such approximations are not available inequalities of the kind of (4.1) or (4.2) may be of practical use.

REFERENCES

[1] H. T. DAVID, "A three-sample Kolmogorov-Smirnov test," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 842–851.

[2] E. F. DRION, "Some distribution-free tests for the difference between two empirical cumulative distribution functions," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 563–574.

[3] B. V. GNEDENKO AND V. S. KOROLYUK, "On the maximum discrepancy between two empirical distributions," *Doklady Akad. Nauk SSSR*, Vol. 80 (1951), pp. 525–528.

[4] J. L. HODGES, JR., "The significance probability of the Smirnov two-sample test," *Arkiv för Matematik*, Vol. 3 (1957), pp. 469–486.

[5] F. J. MASSEY, JR., "The distribution of the maximum deviation between two sample cumulative step functions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 125–128.

[6] N. V. SMIRNOFF, "Sur les écarts de la courbe de distribution empirique," *Mat. Sbornik*, Vol. 6 (48) (1939), pp. 3–26.

[7] N. V. SMIRNOV, "On the estimation of the discrepancy between empirical curves of distribution for two independent samples," *Bull. Math. Univ. Moscou*, Vol. 2. (1939), fasc. 2.

PÓLYA TYPE DISTRIBUTIONS OF CONVOLUTIONS[1]

BY SAMUEL KARLIN AND FRANK PROSCHAN

Stanford University

1. Introduction. The theory of totally positive kernels and Pólya type distributions has been decisively and extensively applied in several domains of mathematics, statistics, economics and mechanics. Totally positive kernels arise naturally in developing procedures for inverting, by differential polynomial operators [7], integral transformations defined in terms of convolution kernels. The theory of Pólya type distributions is fundamental in permitting characterizations of best statistical procedures for decision problems [8] [9] [13]. In clarifying the structure of stochastic processes with continuous path functions we encounter totally positive kernels [11] [12]. Studies in the stability of certain models in mathematical economics frequently use properties of totally positive kernels [10]. The theory of vibrations of certain types of mechanical systems (primarily coupled systems) involves aspects of the theory of totally positive kernels [5].

In this paper, we characterize new classes of totally positive kernels that arise from summing independent random variables and forming related first passage time distributions.

A function $f(x, y)$ of two real variables ranging over linearly ordered one dimensional sets X and Y respectively, is said to be *totally positive of order k* (TP_k) if for all $x_1 < x_2 < \cdots < x_m$, $y_1 < y_2 < \cdots < y_m$, $(x_i \in X; y_j \in Y)$ and all $1 \leqq m \leqq k$,

$$(1)\qquad f\begin{pmatrix} x_1, x_2, \cdots, x_m \\ y_1, y_2, \cdots, y_m \end{pmatrix} = \begin{vmatrix} f(x_1, y_1) & f(x_1, y_2) & \cdots & f(x_1, y_m) \\ f(x_2, y_1) & f(x_2, y_2) & \cdots & f(x_2, y_m) \\ \vdots & \vdots & \ddots & \vdots \\ f(x_m, y_1) & f(x_m, y_2) & \cdots & f(x_m, y_m) \end{vmatrix} \geqq 0.$$

Typically, X is an interval of the real line, or a countable set of discrete values on the real line such as the set of all integers or the set of non-negative integers; similarly for Y. When X or Y is a set of integers, we may use the term "sequence" rather than "function."

A related, weaker property is that of sign regularity. A function $f(x, y)$ is *sign regular of order k*, if for every $x_1 < x_2 < \cdots < x_m$, $y_1 < y_2 < \cdots < y_m$, and $1 \leqq m \leqq k$, the sign of

$$f\begin{pmatrix} x_1, x_2, \cdots, x_m \\ y_1, y_2, \cdots, y_m \end{pmatrix}$$

depends on m alone.

If a TP_k function $f(x, y)$ is a probability density in one of the variables, say x, with respect to a σ-finite measure $\mu(x)$, for each fixed value of y, then $f(x, y)$

Received August 24, 1959; revised April 1, 1960.

[1] This work was supported by the Office of Naval Research under Task NR 047-019.

is said to be *Pólya type of order k* (PT_k). The concepts of PT_1 and PT_2 densities are familiar ones. Every density characterized by a parameter is PT_1; while the PT_2 densities are those having a monotone likelihood ratio [13].

A further specialization occurs if a PT_k kernel may be written as a function $f(x - y)$ of the difference of x and y where x and y traverse the real line; $f(u)$ is then said to be a *Pólya frequency density of order k* (PF_k).

Finally, if the subscript ∞ is written in any of the definitions, then the property in question will be understood to hold for all positive integers.

2. Summary of Results. From Lemma 3 below we trivially obtain the result that if $f_1, f_2, \cdots$ are density functions of non-negative random variables with each f_i a PF_k, then $g(n, x) = f_1 * f_2 * \cdots * f_n(x)$ ($*$ indicates convolution) is PF_k in differences of x for each $n > 0$. One of the key results of this paper is that under the same hypothesis $g(n, x)$ is PT_k in the variables n and x, where n ranges over the positive integers and x traverses the positive real line. That is, total positivity in translation variables (differences of the argument) for each density implies total positivity in the *pair*: the argument and the order of the convolution. (Theorem 1 of Section 4.)

As an easy consequence, we obtain that

$$h(n, x) = P[\sum_{i=1}^{n} X_i \geqq x],$$

where the X_i are independent observations from the corresponding f_i, $i = 1, 2, \cdots$, is TP_k in the variables n and x. The kernel $h(n, x)$ can be interpreted as the probability that first passage into the set $[x, \infty)$ occurs at or before the nth transition where the successive partial sums $S_n = \sum_{i=1}^{n} X_i$, $n = 0, 1, 2, \cdots$ ($S_0 = 0$) describe a discrete time real valued Markov process. If X_i are not identically distributed then the process is not time homogeneous. In this formulation the statement concerning the first passage probability function can be extended to the case of random variables ranging over the whole real line. Thus Theorem 2 of Section 4 asserts that for Pólya frequency densities of a given order, the probability that first passage into the set $[x, \infty)$ for the stochastic process of successive partial sums occurs at the nth transition, is a totally positive function in the variables n, x of the same order. In this framework, Theorem 1 can be deduced from Theorem 2 by employing a suitable limiting argument. Further results of this sort are given in Section 4 and Section 5.

A different kind of characterization is given in Theorem 8 of Section 6. There it is shown that $g(n, x)$, the n-fold convolution of a PF_k density extending over the whole real line, although not possessing the full variation diminishing property of a TP_k function, does possess a restricted variation limiting property. Specifically, $\sum_{i=1}^{m} a_i g(n_i, x)$ has at most $2(m - 1)$ sign changes, where

$$n_1 < n_2 < \cdots < n_m, \; m \leqq (k + 1)/2,$$

and the a_i are real non-zero constants.[2]

[2] The number of sign changes $V(f)$ of a real valued function f is $\sup_{1 \leqq i \leqq m} V(f(x_i))$ where

In Section 7 we establish several smoothening properties possessed by the kernel $f^{(n)}(x)$ (the n-fold convolution of f), when it defines a linear transformation. In particular, we prove that if $f(x)$ is PF_2 and $g(x)$ is convex (concave) then $h(n) = \int f^{(n)}(x)g(x)\,dx$ is convex (concave): This fact is useful in applications.

In Sections 8 and 9 various applications of these results are noted. The inventory problem discussed in Section 8 originally motivated the theoretical results of the present paper; it is exposed here to illustrate the kind of applications made available by exploiting the theorems of Section 4. It is possible to show with the aid of Theorem 1 that the objective function of the inventory problem is concave, so that its maximization becomes a relatively easy task and can be reduced to a rather standard non-linear programming calculation.

In Section 9, a number of totally positive functions are constructed by forming successive convolutions of Pólya frequency densities and then applying Theorem 1. As an illustration of the theory we obtain that $g(n, x) = (x - A_n)^{K_n}$ for $x > A_n$ and 0 for $x \leqq A_n$ is TP_∞ in x and n, provided A_n is any increasing function of n and K_n is any strictly increasing integer-valued function of n.

In a subsequent publication, Karlin will indicate other generalizations and applications of the results of this paper to the theory of stochastic processes and orthogonal polynomials. For example, we will extend the results from a discrete time formulation corresponding to integer convolutions to a continuous time stochastic process structure. In this framework the present theory bears a close relationship to some recent studies of Karlin and McGregor [11] concerned with totally positive kernels and diffusion processes. We will also develop further the connections of total positivity and absorption and recurrence probabilities for the state variable of certain kinds of stochastic processes.

In [15], Proschan has discussed in detail the inventory model described in Section 8 with applications to some concrete examples. Theorem 1 plays a crucial role in this study.

3. Preliminaries. Many of the structural properties of TP_k functions are deducible from the following identity, which appears in [14], p. 48, problem 68:

LEMMA 1: *If $r(x, w) = \int p(x, t)q(t, w)\,d\sigma(t)$ and the integral converges absolutely, then*

$$r\begin{pmatrix} x_1, x_2, \cdots, x_k \\ w_1, w_2, \cdots, w_k \end{pmatrix} = \iint_{t_1<t_2<\cdots<t_k} \cdots \int \tag{2}$$
$$\cdot p\begin{pmatrix} x_1, x_2, \cdots, x_k \\ t_1, t_2, \cdots, t_k \end{pmatrix} q\begin{pmatrix} t_1, t_2, \cdots, t_k \\ w_1, w_2, \cdots, w_k \end{pmatrix} d\sigma(t_1)\,d\sigma(t_2). \quad \cdots d\sigma(t_k).$$

In particular, we secure from Lemma 1, the following useful result:

LEMMA 2: *If $f(x, t)$ is TP_m and $g(t, w)$ is TP_n, then $h(x, w) = \int f(x, t)\, g(t, w)d\sigma(t)$ is $TP_{\min(m,n)}$* provided σ (t) is a regular σ finite measure.

$V(f(x_i))$ is the number of sign changes of the sequence $f(x_1), f(x_2), \cdots, f(x_m)$ with x_i chosen arbitrarily from the domain of definition of f and arranged so that $x_1 < x_2 < \cdots < x_m$ and m any positive integer.

We shall exploit this result principally in the case when f and g are Pólya frequency densities: Therefore,

LEMMA 3: *If $f(x)$ is PF_m and $g(x)$ is PF_n, then $h(x) = \int f(x - t)\, g(t)\, dt$ is $PF_{\min(m,n)}$.*

An important feature of totally positive functions is their variation diminishing property: If $f(x, w)$ is TP_k and $g(w)$ changes sign $j \leqq k - 1$ times, then $h(x) = \int f(x, w)\, g(w)\, d\sigma(w)$ changes sign at most j times; moreover, if $h(x)$ actually changes sign j times, then it must change sign in the same order as $g(w)$ as x and w traverse the real line from left to right [8] [9]. This distinctive property underlies many of the applications mentioned above. The variation diminishing property is essentially equivalent to the determinantal inequalities (1).

4. Convolution of Non-Negative Random Variables. We first prove

THEOREM 1: *Let $f_1, f_2, \cdots$ be any sequence of densities of non-negative random variables, with each f_i a PF_k. Then the n-fold convolution $g(n, x) = f_1 * f_2 * \cdots * f_n(x)$ is PT_k in the variables n and x, where n ranges over 1, 2, $\cdots$ and x traverses the positive real line.*

PROOF: The proof employs induction. First note that $g(n, x)$ is PT_1 since $g(n, x) \geqq 0$ for each real x and each positive integer n.

Assume now that for *every* sequence of densities satisfying the hypothesis, the corresponding n-fold convolution has been proven PT_{r-1} for $r \leqq k$. We prove that this implies $g(n, x)$ is PT_r.

(a) First consider the case $n_1 = 1$. Given $1 < n_2 < n_3 < \cdots < n_r$, $0 \leqq x_1 < x_2 < \cdots < x_r$, we may write

$$(3)\qquad g\begin{pmatrix}1, n_2, \cdots, n_r\\ x_1, x_2, \cdots, x_r\end{pmatrix} = \sum_{\nu=1}^{r} (-1)^{\nu-1} f_1(x_\nu)\, g\begin{pmatrix}n_2, n_3, \cdots & & & , n_r\\ x_1, \cdots, x_{\nu-1}, & x_{\nu+1}, & \cdots, & x_r\end{pmatrix}$$

simply by expanding the determinant on the left by its first row. Next note that for $n = 2, 3, \cdots$ and $x \geqq 0$,

$$(4)\qquad g(n, x) = \int g_1(n - 1, \xi) f_1(x - \xi)\, d\xi,$$

where $g_1(n - 1, \xi)$ is defined as $f_2 * f_3 * \cdots * f_n(\xi)$. Applying (2) in (4), we may write

$$(5)\qquad \begin{aligned} g\begin{pmatrix}n_2, n_3, \cdots & & & , n_r\\ x_1, \cdots, x_{\nu-1}, & x_{\nu+1}, & \cdots, & x_r\end{pmatrix} &= \iint_{0 \leqq \xi_1 < \xi_2 < \cdots < \xi_{r-1}} \cdots \int \\ &\cdot g_1\begin{pmatrix}n_2 - 1, n_3 - 1, \cdots, n_r - 1\\ \xi_1, \xi_2, \cdots \qquad , \xi_{r-1}\end{pmatrix} f_1\begin{pmatrix}x_1, \cdots, x_{\nu-1}, x_{\nu+1}, \cdots, x_r\\ \xi_1, \xi_2, \qquad \cdots \qquad , \xi_{r-1}\end{pmatrix}. \\ &\qquad d\xi_1\, d\xi_2 \cdots d\xi_{r-1}. \end{aligned}$$

Inserting (5) into (3), we get immediately,

$$g\begin{pmatrix}1, n_2, \cdots, n_r\\ x_1, x_2, \cdots, x_r\end{pmatrix} = \iint_{0<\xi_1<\xi_2<\cdots<\xi_{r-1}} \cdots \int g_1\begin{pmatrix}n_2-1, n_3-1, \cdots, n_r-1\\ \xi_1, \xi_2, \cdots \qquad , \xi_{r-1}\end{pmatrix}.$$

(6)

$$\sum_{\nu=1}^{r} (-1)^{\nu-1} f_1(x_\nu) f_1\begin{pmatrix}x_1, \cdots, x_{\nu-1}, x_{\nu+1}, \cdots, x_r\\ \xi_1, \xi_2, \cdots \qquad , \xi_{r-1}\end{pmatrix} d\xi_1\, d\xi_2 \cdots d\xi_{r-1} =$$

$$\iint_{0\le\xi_1<\xi_2<\cdots<\xi_{r-1}} \cdots \int g_1\begin{pmatrix}n_2-1, n_3-1, \cdots, n_r-1\\ \xi_1, \xi_2, \cdots \qquad , \xi_{r-1}\end{pmatrix}$$

$$\cdot f_1\begin{pmatrix}x_1, x_2, \cdots \quad , x_r\\ 0, \xi_1, \xi_2, \cdots, \xi_{r-1}\end{pmatrix} d\xi_1\, d\xi_2 \cdots d\xi_{r-1}.$$

But $g_1\begin{pmatrix}n_2-1, n_3-1, \cdots, n_r-1\\ \xi_1 \quad , \xi_2 \quad , \cdots, \xi_{r-1}\end{pmatrix} \geqq 0$ by the induction assumption, while $f_1\begin{pmatrix}x_1, x_2, x_3, \cdots, x_r\\ 0, \xi_1, \xi_2, \cdots, \xi_{r-1}\end{pmatrix} \geqq 0$ since f_1 is PF_k by the hypothesis of the theorem since $0 \leqq \xi_1 < \xi_2 < \cdots < \xi_{r-1}$. Hence

$$g\begin{pmatrix}1, n_2, \cdots, n_r\\ x_1, x_2, \cdots, x_r\end{pmatrix} \geqq 0. \tag{7}$$

(b) Now suppose $n_1 > 1$. Then for any $n_1 < n_2 < \cdots < n_k$ and $0 \leqq x_1 < x_2 < \cdots < x_k$, we may write, using (2) and (4):

$$g\begin{pmatrix}n_1, n_2, \cdots, n_r\\ x_1, x_2, \cdots, x_r\end{pmatrix} = \iint_{\xi_1<\xi_2<\cdots<\xi_r} \cdots \int g_1\begin{pmatrix}n_1-1, n_2-1, \cdots, n_r-1\\ \xi_1 \quad , \xi_2 \quad , \cdots, \xi_r\end{pmatrix}$$

(8)

$$\cdot f_1\begin{pmatrix}x_1, x_2, \cdots, x_r\\ \xi_1, \xi_2, \cdots, \xi_r\end{pmatrix} d\xi_1\, d\xi_2 \cdots d\xi_r.$$

From (8) we see that for every sequence of densities satisfying the hypothesis, the corresponding functions g_1, g satisfy

$$g_1\begin{pmatrix}n_1-1, n_2-1, \cdots, n_r-1\\ \xi_1 \quad , \xi_2 \quad , \cdots, \xi_r\end{pmatrix} \geqq 0 \Rightarrow g\begin{pmatrix}n_1, n_2, \cdots, n_r\\ x_1, x_2, \cdots, x_r\end{pmatrix} \geqq 0. \tag{9}$$

Using (7) and (9), it follows by induction that $g\begin{pmatrix}n_1, n_2, \cdots, n_r\\ x_1, x_2, \cdots, x_r\end{pmatrix} \geqq 0$.

Since $g(n, x)$ has thereby been proven PT_r, we have established the validity of the induction step, and the theorem follows.

It is important to emphasize the distinction between Lemma 3 and Theorem 1. Under the hypothesis of Theorem 1, Lemma 3 states that for each fixed positive integer n, $g(n, x)$ is PT_k in differences of x, while Theorem 1 states that $g(n, x)$ is PT_k in the variables n and x.

Will Theorem 1 hold if the random variables are not restricted to be non-negative? In general, the answer is no, as the following example shows.

EXAMPLE: Let $f_1(x) = f_2(x) = \cdots = (1/\sqrt{2\pi})\, e^{-x^2/2}$, a PF_∞. Then

$$g(n, x) = f^{(n)}(x) = (1/\sqrt{2\pi n})\, e^{-x^2/2n}.$$

For $1 \leqq n_1 < n_2$, $x_1 < x_2$, the second order determinant is positive for $0 \leqq x_1 < x_2$ and negative for $x_1 < x_2 \leqq 0$. Thus $g(n, x)$ is *not* PT_2.

However a generalization of Theorem 1 to the case of random variables ranging over the whole real line *is* possible, as developed in Theorem 2 below. In the more general case, total positivity holds, not for the n-fold convolution, but rather for the first passage time probabilities of the partial sum process.

THEOREM 2: *Let $f_1, f_2, \cdots$ be any sequence of PT_k densities of random variables $X_1, X_2, \cdots$ respectively, which are not necessarily non-negative. Consider the first passage probability for x positive:*

$$h(n, x) = P\left[\sum_{i=1}^{n} X_i \geqq x; \quad \sum_{i=1}^{j} X_i < x, j = 1, 2, \cdots, n-1\right] \qquad \text{for } n = 1, 2, \cdots.$$

Then $h(n, x)$ is TP_k, where n ranges over $1, 2, \cdots$ and x traverses the positive axis.

PROOF: The proof proceeds in a similar fashion to that of Theorem 1. We employ induction. First we note that $h(n, x)$ is TP_1 since $h(n, x) \geqq 0$ by its very meaning.

Assume now that for every sequence of densities satisfying the hypothesis, the associated first passage time probability function is TP_{r-1} for $r \leqq k$. We shall prove that this implies $h(n, x)$ is TP_r. From this the conclusion of the theorem will follow.

We clearly have for x positive that

$$h(n, x) = \begin{cases} \displaystyle\int_x^\infty f(\xi)\, d\xi & \text{for } n = 1 \\ \displaystyle\int_0^\infty f(x - \xi) h_1(n-1, \xi)\, d\xi & \text{for } n \geqq 2 \end{cases} \tag{10}$$

where

$$h_1(n-1, \xi) = P\left[\sum_{i=2}^{n} X_i \geqq \xi; \quad \sum_{i=2}^{j} X_i < \xi, j = 2, 3, \cdots, n-1\right].$$

We consider first the case $n_1 = 1$. Given $1 < n_2 < n_3 < \cdots < n_r$, $x_1 < x_2 < \cdots < x_r$, we may write, using (10),

$$h\begin{pmatrix} 1, n_2, n_3, \cdots, n_r \\ x_1, x_2, x_3, \cdots, x_r \end{pmatrix} = \begin{vmatrix} \int_0^\infty f_1(x_1 + \xi)\, d\xi & \int_0^\infty f_1(x_2 + \xi)\, d\xi & \cdots & \int_0^\infty f_1(x_r + \xi)\, d\xi \\ h(n_2, x_1) & h(n_2, x_2) & \cdots & h(n_2, x_r) \\ \vdots & \vdots & \ddots & \vdots \\ h(n_r, x_1) & h(n_r, x_2) & \cdots & h(n_r, x_r) \end{vmatrix}$$

$$= \sum_{\nu=1}^{r} (-1)^{\nu-1} \int_0^\infty f_1(x_\nu + \xi) h\begin{pmatrix} n_2, n_3, \cdots & & , n_r \\ x_1, \cdots, x_{\nu-1}, x_{\nu+1}, \cdots & & , x_r \end{pmatrix} d\xi.$$

Now using (10) and (2), we obtain

$$h\begin{pmatrix} n_2, n_3, \cdots & & , n_r \\ x_1, \cdots, x_{\nu-1}, x_{\nu+1}, \cdots, x_r \end{pmatrix} = \iint_{0 \le \xi_1 < \xi_2 < \cdots < \xi_{r-1}} \cdots \int$$

$$(11) \qquad \cdot f_1\begin{pmatrix} x_1, \cdots, x_{\nu-1}, x_{\nu+1}, \cdots, x_r \\ \xi_1, \xi_2, \cdots \qquad , \xi_{r-1} \end{pmatrix} h_1\begin{pmatrix} n_2 - 1, n_3 - 1, \cdots, n_r - 1 \\ \xi_1, \xi_2, \cdots \qquad , \xi_{r-1} \end{pmatrix}.$$

$$d\xi_1\, d\xi_2 \cdots d\xi_{r-1}.$$

Inserting (11) in the equation above and replacing $-\xi$ by ξ, gives

$$h\begin{pmatrix} 1, n_2, n_3, \cdots, n_r \\ x_1, x_2, x_3, \cdots, x_r \end{pmatrix} = \iint_{\xi < 0 \le \xi_1 < \xi_2 < \cdots < \xi_{r-1}} \cdots \int$$

$$\cdot h_1\begin{pmatrix} n_2 - 1, n_3 - 1, \cdots, n_r - 1 \\ \xi_1, \xi_2, \cdots \qquad , \xi_{r-1} \end{pmatrix} \sum_{\nu=1}^{r} (-1)^{\nu-1} f_1(x_\nu - \xi).$$

$$f_1\begin{pmatrix} x_1, \cdots, x_{\nu-1}, x_{\nu+1}, \cdots, x_r \\ \xi_1, \xi_2, \cdots \qquad , \xi_{r-1} \end{pmatrix} d\xi\, d\xi_1\, d\xi_2 \cdots d\xi_{r-1}$$

$$= \iint_{\xi < 0 \le \xi_1 < \xi_2 < \cdots < \xi_{r-1}} \cdots \int h_1\begin{pmatrix} n_2 - 1, n_3 - 1, \cdots, n_r - 1 \\ \xi_1, \xi_2, \cdots \qquad , \xi_{r-1} \end{pmatrix}$$

$$\cdot f_1\begin{pmatrix} x_1, x_2, \cdots, x_r \\ \xi, \xi_1, \xi_2, \cdots, \xi_{r-1} \end{pmatrix} d\xi\, d\xi_1\, d\xi_2 \cdots d\xi_{r-1}.$$

But $h_1\begin{pmatrix} n_2 - 1, n_3 - 1, \cdots, n_r - 1 \\ \xi_1 \quad , \xi_2 \quad , \cdots, \xi_{r-1} \end{pmatrix} \geqq 0$ by the inductive assumption, while $f_1\begin{pmatrix} x_1, x_2, \cdots \quad , x_r \\ \xi, \xi_1, \xi_2, \cdots, \xi_{r-1} \end{pmatrix} \geqq 0$ since $\xi < \xi_1 < \xi_2 < \cdots < \xi_{r-1}$, $x_1 < x_2 < \cdots < x_r$, and $f_1(x)$ is PF_k by hypothesis. Hence $h\begin{pmatrix} 1, n_2, n_3, \cdots, n_r \\ x_1, x_2, x_3, \cdots, x_r \end{pmatrix} \geqq 0$.

The remainder of the proof parallels the corresponding portion of the proof of Theorem 1; simply replace g by h.

It is appropriate to compare Theorem 1 and Theorem 2. For this purpose we sketch an argument which shows that Theorem 1 is actually a limiting case of Theorem 2. A careful examination of the preceding argument reveals that in the case of non-negative PF_k random variables, the probability of first passage at time n into any positive interval, not only the interval $[x, \infty]$, is TP_k. In view of this fact we shrink the interval to a point, and it readily follows that the first passage time probability converges to the density corresponding to the n-fold convolution. Since total positivity is preserved under this limiting operation, Theorem 1 follows.

We now develop a series of consequences of Theorems 1 and 2. Let $F_i(x)$ be the cumulative distribution function corresponding to $f_i(x)$, $i = 1, 2, \cdots$. Then as a direct corollary of Theorem 2, we have

THEOREM 3: *Under the assumptions of Theorem* 1,

$$h(n, x) = F_1 * F_2 * \cdots * F_{n-1}(x) - F_1 * F_2 * \cdots * F_n(x) \quad \text{is} \quad TP_k,$$

where n ranges over $1, 2, \cdots$ *and* $x > 0$. *In particular, if* $f_i = f$, $i = 1, 2, \cdots$, *then* $h(n, x) = F^{(n-1)}(x) - F^{(n)}(x)$ *is* TP_k.

PROOF: Simply note that

$$h(n, x) = P[\textstyle\sum_{i=1}^{n} X_i \geq x; \quad \sum_{i=1}^{n-1} X_i < x]$$

since the random variables are non-negative.

Actually we can say more about $h(n, x)$ in the situation where the X_i are non-negative, independent, and identically distributed random variables; Theorem 4 asserts that for each fixed $x > 0$, $h(n + m, x)$ is sign regular in the variables $n \geq 0$ and $m \geq 0$.

THEOREM 4: *Suppose* $f(x)$ *is* PF_k *with* $f(x) = 0$ *for* $x < 0$. *We define* $h(n, x)$ *by* $h(n, x) = F^{(n-1)}(x) - F^{(n)}(x)$ *for* $n = 1, 2, \cdots$; $x \geq 0$, *and for fixed* $x \geq 0$ *we define*

$$c(n) = \begin{cases} h(n, x) & \text{for } n \geq 1 \\ 0 & \text{for } n \leq 0. \end{cases}$$

Then $c(n + m)$ *is sign regular of order* k *in* $n \geq 1$ *and* $m \geq 1$; *moreover, for* $1 \leq n_1 < n_2 < \cdots < n_r$, $1 \leq m_1 < m_2 < \cdots < m_r$, *the sign of*

$$c_+\begin{pmatrix} n_1, n_2, \cdots, n_r \\ m_1, m_2, \cdots, m_r \end{pmatrix} \text{ is } (-1)^{r(r-1)/2}, \text{ where } c_+\begin{pmatrix} n_1, n_2, \cdots, n_r \\ m_1, m_2, \cdots, m_r \end{pmatrix} = \begin{vmatrix} c(n_1 + m_1) & \cdots & c(n_1 + m_r) \\ \vdots & \cdot & \vdots \\ c(n_r + m_1) & \cdots & c(n_r + m_r) \end{vmatrix}.$$

PROOF: For $m \geq 1$ and $n \geq 1$, we have

$$c(n + m) = \int g(m, \xi) h(n, x - \xi) \, d\xi; \tag{12}$$

where $g(m, \xi) = f^{(m)}(\xi)$. (12) simply states that if the partial sum first exceeds x at the $n + m$th stage, then this can occur by having the mth partial sum equal to some non-negative $\xi < x$, while the partial sum starting with the $m + 1$st variable first exceeds $x - \xi$ at the nth stage. From (12) and (2), we get, for $1 \leq n_1 < n_2 < \cdots < n_r$, $1 \leq m_1 < m_2 < \cdots < m_r$, $r \leq k$,

$$\begin{aligned} c_+\begin{pmatrix} n_1, n_2, \cdots, n_r \\ m_1, m_2, \cdots, m_r \end{pmatrix} &= \underset{0 \leq \xi_1 < \xi_2 < \cdots < \xi_r < x}{\iint \cdots \int} g\begin{pmatrix} m_1, m_2, \cdots, m_r \\ \xi_1, \xi_2, \cdots, \xi_r \end{pmatrix} \\ &\quad \cdot h\begin{pmatrix} n_1 & , n_2 & , \cdots, n_r \\ x - \xi_1, & x - \xi_2, & \cdots, x - \xi_r \end{pmatrix} d\xi_1, d\xi_2, \cdots, d\xi_r. \end{aligned} \tag{13}$$

By Theorem 1, $g\begin{pmatrix} m_1, m_2, \cdots, m_r \\ \xi_1, \xi_2, \cdots, \xi_r \end{pmatrix} \geq 0$. Since the $x - \xi_1, x - \xi_2, \cdots, x - \xi_r$ are in decreasing order of magnitude it follows, invoking Theorem 3, that

$h\begin{pmatrix} n_1 & , n_2 & , \cdots, n_r \\ x - \xi_1, & x - \xi_2, & \cdots, x - \xi_r \end{pmatrix}$ has the sign $(-1)^{r(r-1)/2}$. Thus

$$c_+\begin{pmatrix} n_1, & n_2, & \cdots, & n_r \\ m_1, & m_2, & \cdots, & m_r \end{pmatrix}$$

has the sign $(-1)^{r(r-1)/2}$ as was to be proved.

We prove below that $c(n)$ has the property of being PF_2 provided $f(x)$ is PF_2 (Theorem 5). This is the property required for the analysis of the inventory model of Section 8. In contrast, this relationship between $c(n)$ and $f(x)$ does not persist beyond the second order.

THEOREM 5: *If $f(x)$ is PF_2 with $f(x) = 0$ for $x < 0$, then $c(n)$ (defined in Theorem 4) is PF_2.*

PROOF: Let $n_1 < n_2$, $m_1 < m_2$. Write

$$c\begin{pmatrix} n_1, & n_2 \\ m_1, & m_2 \end{pmatrix} = \begin{vmatrix} c(n_1 - m_1) & c(n_1 - m_2) \\ c(n_2 - m_1) & c(n_2 - m_2) \end{vmatrix}.$$

(a) If $n_1 \leqq m_2$, then $c(n_1 - m_2) = 0$, so that $c\begin{pmatrix} n_1, & n_2 \\ m_1, & m_2 \end{pmatrix} \geqq 0$.

(b) If $n_1 > m_2$, we must have $m_1 < m_2 < n_1 < n_2$. Hence

$$c\begin{pmatrix} n_1, & n_2 \\ m_1, & m_2 \end{pmatrix} = \begin{vmatrix} h(n_1 - m_1, x) & h(n_1 - m_2, x) \\ h(n_2 - m_1, x) & h(n_2 - m_2, x) \end{vmatrix}$$

$$= \int \begin{vmatrix} h(n_1 - m_2, \xi) & h(n_1 - m_2, x) \\ h(n_2 - m_2, \xi) & h(n_2 - m_2, x) \end{vmatrix} g(m_2 - m_1, x - \xi)\, d\xi.$$

Since $\xi < x$, $n_1 - m_2 < n_2 - m_2$, and $h(n, x)$ is TP_2 by Theorem 3, then $c\begin{pmatrix} n_1, & n_2 \\ m_1, & m_2 \end{pmatrix} \geqq 0$ and the proof is finished.

5. Compound Distributions. As an easy corollary of Theorem 1, we have corresponding determinantal properties for compound distributions composed from PT_k densities. Specifically:

THEOREM 6: *Let $X_i \geqq 0$ be distributed with density $f_i(x)$, a PF_k, $i = 1, 2, \cdots$. Define $S_n = \sum_{i=1}^{N} X_i$, where N is a random variable independent of $X_1, X_2, \cdots$, with density $d(n, \mu)$, where μ is a parameter, and $d(n, \mu)$ is PT_k in the variables n and μ. Then $r(x, \mu)$, the probability density for S_N, is PT_k in the variables $x > 0$ and μ.*

PROOF: $r(x, \mu) = \sum_{n=1}^{\infty} P[N = n]\, f_1 * f_2 * \cdots * f_n(x) = \sum_{n=1}^{\infty} d(n, \mu)\, g(n, x)$. By Theorem 1, $g(n, x)$ is PT_k. Applying Lemma 2 we conclude that $r(x, \mu)$ is also PT_k.

In a similar fashion, we may study transforms of $g(n, x)$ in the variable x; the proof is as in Theorem 6.

THEOREM 7: *In addition to the hypothesis of Theorem 1 assume that $\varphi(x, s)$ is a PT_k function. Then $\phi(n, s) = \int g(n, x)\varphi(x, s)\, dx$ is PT_k.*

As an illustration, let $\varphi(x, s) = e^{sx}$, $-\infty < s \leqq 0$, so that $\varphi(x, s)$ is PT_∞. From Theorem 7, we have that $\phi(n, s)$ is PT_k in the variables n and s. But $\phi(n, s)$ is the Laplace transform of the convolution of n densities and so we have $\phi(n, s) = \phi_1(s)\phi_2(s) \cdots \phi_n(s)$, where $\phi_i(s) = \int f_i(x)e^{sx}\,dx$, $i = 1, 2, \cdots$. In particular, we obtain the interesting set of inequalities:

$$\begin{vmatrix} \phi_1(s_1) & \phi_1(s_1)\phi_2(s_1) & \cdots & \phi_1(s_1)\phi_2(s_1) & \cdots & \phi_m(s_1) \\ \phi_1(s_2) & \phi_1(s_2)\phi_2(s_2) & \cdots & \phi_1(s_2)\phi_2(s_2) & \cdots & \phi_m(s_2) \\ \vdots & \vdots & & \vdots & & \\ \phi_1(s_m) & \phi_1(s_m)\phi_2(s_m) & \cdots & \phi_1(s_m)\phi_2(s_m) & \cdots & \phi_m(s_m) \end{vmatrix} \geqq 0$$

where $s_1 < s_2 < \cdots < s_m \leqq 0$, $m \leqq k$, $\phi_i(s) = \int f_i(x)e^{sx}\,dx$, and $f_i(x)$ is a PF_k density with $f_i(x) = 0$ for $x < 0$, $i = 1, 2, \cdots, k$.

6. Convolution of Random Variables Ranging Over the Real Line. We have seen on the basis of the example following Theorem 1, that the n-fold convolution $g(n, x)$ of a PF_k density whose possible values extend over the whole real line, is not necessarily PT_k. Thus, in generalizing Theorem 1 to densities whose possible values extend throughout the real line, it was necessary to formulate the problem in terms of first passage probabilities rather than n-fold convolutions. However, the question remains: what smoothening properties are possessed by the n-fold convolution of a PF_k density, which has possible values ranging over the full real line. We can answer this query in terms of a weakened version of the variation diminishing property possessed by totally positive functions. Recall that if $p(x, w)$ is TP_k and $q(w)$ changes sign $j \leqq k - 1$ times, then

$$r(x) = \int p(x, w)q(w)\,dF(w)$$

changes sign at most j times; moreover, if $r(x)$ actually changes sign j times, then it must change sign in the same order as does $q(w)$ [9]. This variation diminishing property may be compared with the following result.

THEOREM 8: *Let $f(x)$ be a continuous PF_k, with $f(x)$ not necessarily 0 for $x < 0$. Let $r_m(x) = \sum_{i=1}^m a_i g(n_i, x)$, where $n_1 < n_2 < \cdots < n_m$, $m \leqq (k + 1)/2$, and the a_i are real non-zero constants. Then $r_m(x)$ has $\leqq 2(m - 1)$ sign changes.*

PROOF: We proceed by induction. The theorem trivially holds for $m = 1$.

Assume the theorem holds for the case of a sum consisting of $m_0 - 1$ terms, where $m_0 \leqq (k + 1)/2$. Write

$$r_{m_0}(x) = \sum_{i=1}^{m_0} a_i g(n_i, x) = \sum_{i=2}^{m_0} a_i \int g(n_i - n_1, \theta)g(n_1, x - \theta)\,d\theta$$
$$+ a_1 \lim_{R\to\infty} \int g_R(0, \theta)g(n_1, x - \theta)\,d\theta,$$

where

$$g_R(0, \theta) = \begin{cases} R & \text{for } 0 \leqq \theta \leqq \dfrac{1}{R} \\ 0 & \text{otherwise.} \end{cases}$$

Factoring, we get

$$(14) \qquad r_{m_0}(x) = \lim_{R\to\infty} \int \left\{ \sum_{i=2}^{m_0} a_i g(n_i - n_1, \theta) + a_1 g_R(0, \theta) \right\} g(n_1, x - \theta)\, d\theta.$$

By the inductive hypothesis, $\sum_{i=2}^{m_0} a_i g(n_i - n_1, \theta)$ has at most $2(m_0 - 2)$ sign changes as a function of θ. With R sufficiently large, $a_1 g_R(0, \theta)$ can introduce at most 2 additional sign changes. Thus for sufficiently large R,

$$\sum_{i=2}^{m_0} a_i g(n_i - n_1, \theta) + a_1 g_R(0, \theta)$$

has at most $2(m_0 - 1)$ sign changes. Since $g(n_1, x - \theta)$ is a PF_k, and therefore, variation diminishing, we obtain that the integral of (14) possesses at most $2(m_0 - 1)$ sign changes as a function of x. Taking the limit as $R \to \infty$, the number of sign changes cannot increase, and thus the number of sign changes of $r_{m_0}(x)$ is $\leqq 2(m_0 - 1)$.

Applying induction, we conclude that the theorem holds for $m = 1, 2, \cdots, (k + 1)/2$ and the proof is finished.

7. Preserving Convexity and Concavity. Let $X_i \geqq 0$ be independent random variables distributed according to $f(x)$, a PF_k. We now describe some further smoothening properties possessed by the transformation which maps functions into sequence, viz.

$$h(n) = \int f^{(n)}(x) g(x)\, dx \qquad \text{for } n = 1, 2, \cdots.$$

We show first that the property of convexity is preserved under this transformation. Explicitly, we prove that convexity in $g(x)$ is carried over into convexity in $h(n)$. This will be demonstrated not only for the ordinary notion of convexity, but for a type of convexity of higher order, which notion is made precise below. Similar results hold for concavity.

Assume $f(x)$ is PF_3 and $g(x)$ is convex (of order 2). Let $\mu_i = \int x^i f(x)\, dx$, $i = 1, 2, \cdots$ represent the moments of X. Note that for arbitrary real constants a_0 and a_1,

$$\int \{g(x) - [(a_0/\mu_1)x + a_1]\} f^{(n)}(x)\, dx = h(n) - (a_0 n + a_1).$$

Since $g(x)$ is convex, then $g(x) - [(a_0/\mu_1)x + a_1]$ has at most 2 changes of sign and if 2 changes of sign actually occur, they occur in the order $+ - +$ as x traverses the real axis from $-\infty$ to $+\infty$. Since f is PF_3, then by Theorem 1, $f^{(n)}(x)$ is PT_3 in the variables n and x.

By the variation diminishing property of Pólya type functions, we infer that $h(n) - (a_0 n + a_1)$ will have at most 2 changes of sign. Moreover, if $h(n) - (a_0 n + a_1)$ has exactly 2 changes of sign, then these will occur in the same order as those of $g(x) - [(a_0/\mu_1)x + a_1]$, namely $+ - +$. Since a_0, a_1 are arbitrary, we easily infer that $h(n)$ is a convex function of n.

In a similar fashion we can show that higher order convexity is preserved under this transformation as follows: A function $g(x)$ is said to be convex of order r if for an arbitrary polynomial $p(x) = a_0x^{r-1} + a_1x^{r-2} + \cdots + a_{r-1}$ of degree $r - 1$, $g(x) - p(x)$ has at most r changes of sign, and if r changes of sign actually occur, they occur in the order $+ - + \cdots$.

Assume that $f(x)$ is PF_{r+1} and $g(x)$ is convex of order r. Note that $\int x^k f^{(n)}(x)\, dx = E(X_1 + \cdots + X_n)^k = \mu_1^k n^k +$ lower powers of n. It follows immediately that for an arbitrary polynomial $q(n) = a_0n^{r-1} + a_1n^{r-2} + \cdots + a_{r-1}$ of degree $r - 1$, there exists a polynomial $p(x) = b_0x^{r-1} + b_1x^{r-2} + \cdots + b_{r-1}$ of degree $r - 1$ such that $\int p(x)f^{(n)}(x)\, dx = q(n)$, and hence $\int\{g(x) - p(x)\}f^{(n)}(x)\, dx = h(n) - q(n)$ with $a_0b_0 > 0$. Since $f(x)$ is PF_{r+1}, then by Theorem 1, $f^{(n)}(x)$ is PT_{r+1} in the variables n and x and again by the variation diminishing property of Pólya type functions, we obtain that $h(n) - q(n)$ will have no more changes of sign than $g(x) - p(x)$. But $g(x) - p(x)$ has at most r changes in sign since $g(x)$ is convex of order r; and so $h(n) - q(n)$ has at most r changes of sign. Moreover, if $h(n) - q(n)$ actually has r changes of sign, then they will occur in the same order as those of $g(x) - p(x)$, namely $+ - + \cdots$. Thus $h(n)$ is convex of order r since $q(n)$ was an arbitrary polynomial of degree $r - 1$.

Similar results apply to concavity of higher order. A function $g(x)$ is concave of order r if for an arbitrary polynomial $p(x) = a_0x^{r-1} + a_1x^{r-2} + \cdots + a_{r-1}$ of degree $r - 1$, $g(x) - p(x)$ has at most r changes of sign, and if r changes of sign happen then they occur in the order $- + - \cdots$.

An application may be made to the inventory model discussed in [2], p. 227. The probability density of demand for each period is $f(\xi)$, a PF_2. The policy followed is to maintain the stock size at a fixed level S which will be suitably chosen so as to minimize appropriate expected costs, or is determined by a fixed capacity restriction. At the end of each period an order is placed to replenish the stock consumed during that period so that a constant stock level is maintained on the books. Delivery takes place a periods later. The expected cost for a stationary period as a function of the lag is

$$L(a) = \int_0^S h(S - z)f^{(a)}(z)\, dz + \int_S^\infty p(z - S)f^{(a)}(z)\, dz$$

where S is fixed.

Assume now that h and p are convex increasing functions with $h(0) = p(0) = 0$. Then we may write $L(a) = \int r(z)f^{(a)}(z)\, dz$, where

$$r(z) = \begin{cases} h(S - z) & \text{for} \quad 0 \leq z \leq S \\ p(z - S) & \text{for} \quad S < z. \end{cases}$$

Then $r(z)$ is a convex function. Using the preceding results, we conclude that $L(a)$ is a convex function. Thus, if the length of lag should increase, the marginal expected loss increases.

Similar results hold if p and h are concave. Also, if we assume f is PF_{k+1} and p and h are convex (concave) of order k, we may conclude that $L(a)$ is convex (concave) of order k.

8. Application to an Inventory Problem. We wish to determine the initial spare parts kit for a system, which maximizes assurance of no shortage whatsoever during a period of length t, under a budget for spares c_0. We consider only essential components, and assume that a failed component is instantly replaced by a spare, if available. Only spares initially provided may be used for replacement. The system contains d_i operating components of type i, $i = 1, 2, \cdots, k$. The length of life of the jth operating component of the ith type is an independent random variable with PF_k density f_{ij}, $j = 1, 2, \cdots, d_i$. The unit cost of a component of type i is c_i.

Our problem is to find n_i, the number of spares initially stocked of the ith type, $i = 1, 2, \cdots, k$, such that $\prod_{i=1}^k P_i(n_i)$ is maximized subject to

$$\sum_{i=1}^k n_i c_i \leq c_0 \quad \text{and} \quad n_i = 0, 1, 2, \cdots \qquad \text{for} \quad i = 1, 2, \cdots, k,$$

where $P_i(m)$ = probability of experiencing $\leq m$ failures of type i. (See [3], [15] for a detailed discussion of this model and its application to reliability; our present treatment is confined to aspects of the problem relevant to the present paper.)

In [3] and [15], methods are given for computing the solution when each $\ln P_i(m)$ is concave in m, or equivalently, when each $P_i(n - m)$ is a TP_2 sequence in n and m. To show $P_i(n - m)$ is a TP_2 sequence in n and m, we note:

1. $c_{ij}(n)$, the probability of requiring n replacements of operating component i, j, is a PF_2 sequence in n for each fixed i, j by Theorem 5 above.
2. $p_i(n)$, the probability of requiring n replacements of type i, is a PF_2 sequence in n for each i by Lemma 3, since $p_i(n) = c_{i1} * c_{i2} * \cdots * c_{id_i}(n)$.
3. $P_i(n - m)$ is a TP_2 sequence in n, m for each i, since

(a)
$$P_i(n) = \sum_{m=-\infty}^{\infty} p_i(n - m)q(m), \qquad \text{where } q(m)$$
$$= \begin{cases} 1 & \text{for} \quad m = 0, 1, 2, \cdots \\ 0 & \text{otherwise,} \end{cases}$$

(b) $q(m)$ is a PF_∞ sequence.

(c) The convolution of PF_k sequences is PF_k, by Lemma 3.

Thus when the underlying densities for the life of components are PF_2, the methods given in [3] and [15] for obtaining optimal kits are applicable.

9. Generating Totally Positive Functions. In this section we give a series of examples of the above theorems. These theorems are written in terms of real valued random variables but it should be emphasized that all our results are equally valid for integer valued random variables. The underlying densities

are assumed to be the appropriate PF_k sequences. The first few illustrations involve integer valued random variables.

EXAMPLE 1:

(a) Let

$$f(k) = \begin{cases} q & \text{for } k = 0 \\ p & \text{for } k = 1, \\ 0 & \text{for other } k \end{cases} \qquad \text{where } p + q = 1.$$

then $f(k)$ is a PF_∞ sequence by direct verification. Alternately, we may appeal to a classical result of Schoenberg and Edrei which asserts that a sequence is a PF sequence, if and only if its generating function is of the form

$$e^{\gamma s} \prod_1^\infty \{(1 + \alpha_\nu s)/(1 - \beta_\nu s)\}, \gamma \geqq 0, \alpha_\nu \geqq 0, \beta_\nu \geqq 0; \sum \alpha_\nu \text{ and } \sum \beta_\nu$$

convergent. (See p. 305, [6].) Applying Theorem 1, we obtain that the binomial density $g(n, k) = f^{(n)}(k) = \binom{n}{k} p^k q^{n-k}$ is PT_∞. It follows that $\binom{n}{k}$ is TP_∞ in the variables n and k.

A direct proof, in this case, is easy. For some of our further examples the result is less apparent

(b) Let

$$f_i(k) = \begin{cases} 1/(1 + p^i) & \text{for } k = 0 \\ p^i/(1 + p^i) & \text{for } k = 1 \\ 0 & \text{for other } k, \end{cases}$$

$i = 1, 2, \cdots$. As pointed out in (a) above, each $\{f_i(k)\}_{k=0,1,\cdots}$ is a PF_∞ sequence. Hence, by Theorem 1, $g(n, k) = f_1 * f_2 * \cdots * f_n(k)$ is PT_∞. But $g(n, k)$ is simply the coefficient of s^k in the generating function $\prod_{i=1}^n [(1 + p^i s)/(1 + p^i)]$ of the n-fold convolution. Using the Gauss identity

$$\prod_{i=1}^n (1 + p^i s) = \sum_{\nu=0}^n \begin{bmatrix} n \\ \nu \end{bmatrix} p^{(\nu^2+\nu)/2} s^\nu,$$

where, by definition,

$$\begin{bmatrix} n \\ \nu \end{bmatrix} = \{(1 - p^n)(1 - p^{n-1}) \cdots (1 - p^{n-\nu+1})/(1 - p^\nu)(1 - p^{\nu-1}) \cdots (1 - p) \qquad \text{for } \nu \leqq n,$$

we find that the coefficient of s^k in $\prod_{i=1}^n [(1 + p^i s)/(1 + p^i)]$ is $\begin{bmatrix} n \\ k \end{bmatrix} p^{(k^2+k)} \Big/ \prod_{i=1}^n (1 + p^i)$. Since $p^{(k^2+k)/2}$ is a function of k alone while $\prod_{i=1}^n (1 + p^i)$ is a function of n alone, we conclude that $\begin{bmatrix} n \\ k \end{bmatrix}$ is TP_∞. Note that $\begin{bmatrix} n \\ k \end{bmatrix}$ is a type of generalization of the binomial coefficient $\binom{n}{k}$ since for $p \to 1$, $\begin{bmatrix} n \\ k \end{bmatrix} \to \binom{n}{k}$.

(c) Let $f(k) = q^k p$, $p + q = 1$, $k = 0, 1, 2, \cdots$; $f(k)$, the geometric density, is the probability that the first success in a sequence of Bernoulli trials occurs following k successive failures. The corresponding generating function is $p/(1 - qs)$. By [1], p. 305, $f(k)$ is PF_∞. Now

$$g(n, k) = f^{(n)}(k) = \binom{n + k - 1}{k} p^n q^k,$$

so that $g(n, k)$ represents the probability that the nth success occurs at trial $n + k$ in the sequence of Bernoulli trials. By Theorem 1, $g(n, k)$ is PT_∞. Since p^n is a function of n only, while q^k is a function of k only, we obtain that $\binom{n + k - 1}{k}$ is TP_∞.

(d) Next, let $f_i(k) = q^{ik}(1 - q^i)$, $k = 0, 1, 2, \cdots$, $i = 1, 2, \cdots$. As noted in (c), each $\{f_i(k)\}_{k=0,1\cdots}$ is a PF_∞ sequence. Hence

$$g(n, k) = f_1 * f_2 * \cdots * f_n(k)$$

is PT_∞ by Theorem 1. But $g(n, k)$ is simply the coefficient of s^k in the generating function $\prod_{i=1}^n [(1 - q^i)/(1 - q^i s)]$. Using the Heine hypergeometric relation, [6], p. 8,

$$1 \Big/ \prod_{i=1}^{n} (1 - q^i s) = \sum_{k=1}^{\infty} \frac{[n + 1][n + 2] \cdots [n + k]}{[k][k - 1] \cdots [1]} s^k,$$

where the symbol $[m]$ is defined equal to $[(1 - q^m)/(1 - q)]$. We find that the coefficient of s^k in the generating function is

$$\left\{ \prod_{i=1}^{n} (1 - q^i) \right\} \frac{[n + 1][n + 2] \cdots [n + k]}{[k][k - 1] \cdots [1]}.$$

Since $\prod_{i=1}^n (1 - q^i)$ is a function of n alone, we obtain that

$$\frac{[n + 1][n + 2] \cdots [n + k]}{[k][k - 1] \cdots [1]} \text{ is } TP_\infty.$$

Next we consider an example of the application of Theorem 1 to continuous densities

(e) Let

$$f_i(x) = \begin{cases} (x - a_i)^{k_i - 1} e^{-(x - a_i)}/\Gamma(k_i) & \text{for } x \geq a_i \\ 0 & \text{for } x < a_i, \end{cases}$$

where k_i is a positive integer, $a_i \geq 0$, $i = 1, 2, \cdots$; thus $f_i(x)$ is a translated gamma density. Then the characteristic function of $f_i(x)$,

$$\varphi_i(t) = \int_{a_i}^{\infty} e^{itx}((x - a_i)^{k_i - 1} e^{-(x - a_i)})/\Gamma(k_i)\, dx = e^{ita_i}/(1 - it)^{k_i}.$$

Defining $g(n, x) = f_1 * f_2 * \cdots * f_n(x)$, we have for its characteristic function $\exp [it\sum_{i=1}^n a_i]1/(1 - it)^{\sum_{i=1}^n k_i}$; and consequently

$$g(n, x) = \begin{cases} ((x - A_n)^{K_n-1}e^{-(x-A_n)})/\Gamma(K_n) & \text{for} \quad x \geqq A_n \\ 0 & \text{for} \quad x < A_n, \end{cases}$$

where $A_n = \sum_{i=1}^n a_i$ and $K_n = \sum_{i=1}^n k_i$. This means that $g(n, x)$ is also a translated gamma with parameters corresponding to the sum of the individual parameters.

Since each f_i is PF_∞, we may conclude that $g(n, x)$ is PT_∞ in the variables n and x by Theorem 1, or equivalently, factoring out e^{-x} and $e^{A_n}/\Gamma(K_n)$, that $(x - A_n)^{K_n-1}$ is TP_∞. Note that by appropriate selection of the a_i and the k_i we may achieve for A_n any increasing function of n and $K_n - 1$ may likewise denote any strictly increasing integer-valued function of n.

REFERENCES

[1] Aissen, M., Edrei, A., Schoenberg, I. J., and Whitney, A., "On the generating functions of totally positive sequences," *Proc. Nat. Acad. Sci.*, U.S.A., Vol. 37 (1951), pp. 303–307.

[2] Arrow, K. J., Karlin, S. and Scarf, H., *Studies in the Mathematical Theory of Inventory and Production*, Stanford University Press, Stanford, California, 1958.

[3] Black, G., and Proschan, F., "On optimal redundancy," *Operations Research*, Vol. 7 (1959), pp. 581-588.

[4] Feller, W., *An Introduction to Probability Theory and its Applications*, John Wiley and Sons, New York, 1950.

[5] Gantmacher, F. and Krein, M., *Oscillation Matrices and Kernels and Small Vibrations of Mechanical Systems*, 2nd ed. (in Russian), Moscow, 1950.

[6] Hahn, Wolfgang, "Über Orthogonalpolynome die q-Differenzengleichungen genügen," *Mathematische Nachrichten*, Vol. 2 (1949), pp. 4–34.

[7] Hirschmann, I. I. and Widder, D. V., *The Convolution Transform*, Princeton University Press, Princeton, New Jersey, 1955.

[8] Karlin, S., "Decision theory for Pólya type distributions, Case of Two Actions, I," *Proceedings of the Third Berkeley Symposium on Probability and Statistics*, Vol. I, University of California Press, 1956, pp. 115–129.

[9] Karlin, S., "Pólya type distributions, II," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 281–308.

[10] Karlin, S., *Mathematical Methods and Theory in Games, Programming and Economics* Addison Wesley, Cambridge, Mass., 1959.

[11] Karlin, S. and McGregor, J., "Coincidence probabilities," *Pacific J. Math.*, Vol. 9 (1959), pp. 1141–64.

[12] Karlin, S. and McGregor, J., "Coincident properties of birth and death processes," *Pacific J. Math.*, Vol. 9. (1959), pp. 1109–40.

[13] Karlin, S. and Rubin, H., "The theory of decision procedures for distributions with monotone likelihood ratio," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 272–299.

[14] Pólya, G. and Szegö, G., *Aufgaben und Lehrsatze aus der Analysis*, Vol. I, Julius Springer, Berlin, 1925.

[15] Proschan, F., "Pólya type distributions in renewal theory, with an application to an inventory problem," Thesis in the Statistics Department, Stanford University, May, 1959.

[16] Schoenberg, I. J., "On Pólya frequency functions I, the totally positive functions and their Laplace transforms," *J. d'Analyse Mathématique*, Vol. I (1951), pp. 331–374.

THE POISSON APPROXIMATION TO THE POISSON BINOMIAL DISTRIBUTION[1]

By J. L. Hodges, Jr. and Lucien Le Cam

University of California, Berkeley

1. Introduction. It has been observed empirically that in many situations the number S of events of a specified kind has approximately a Poisson distribution. As examples we may mention the number of telephone calls, accidents, suicides, bacteria, wars, Geiger counts, Supreme Court vacancies, and soldiers killed by the kick of a horse.

Many textbooks in probability content themselves with an explanation of this phenomenon that runs something like this: There is a large number, say n, of events that might occur—for example, there are many telephone subscribers who might place a call during a given minute. The chance, say p, that any specified one of these events will occur (e.g., that a specified telephone subscriber will call), is small. Assuming that the events are independent, S has exactly the binomial distribution, say $\mathcal{B}(n, p)$. If we now let $n \to \infty$ and $p \to 0$, so that $np \to \lambda$ where λ is fixed and $0 < \lambda < \infty$, it is shown that $\mathcal{B}(n, p)$ tends to the Poisson distribution $\mathcal{P}(\lambda)$ with expectation λ.

As was pointed out by von Mises [4], such an explanation is often not satisfactory because the various trials cannot in many applications reasonably be regarded as equally likely to succeed. Let p_i denote the success probability of the ith trial, $i = 1, 2, \cdots, n$. Then S has the distribution sometimes called "Poisson binomial." Starting from this more realistic model von Mises shows that S has in the limit the distribution $\mathcal{P}(\lambda)$, provided $n \to \infty$ and the p_i vary with n in such a way that $\Sigma p_i = \lambda$ is fixed and $\alpha = \max\{p_1, p_2, \cdots, p_n\}$ tends to 0. This result is given in a few textbooks [1], [5].

The limit theorem of von Mises suggests that the Poisson approximation will be reliable provided that n is large, α is small, and λ is moderate. But even these requirements are unnecessarily restrictive, as may be seen from a general approximation theorem of Kolmogorov [2]. When this theorem is applied to our problem, it asserts that there is some constant C, independent of n and the p_i, such that the maximum absolute difference D between the cumulative distributions of S and of $\mathcal{P}(\Sigma p_i)$ satisfies the bound $D \leq C \sqrt[5]{\alpha}$. Thus, the Poisson approximation will be good provided only α is small, whether n is small or large, and whatever value Σp_i may have. It seems to us that this is the type of theorem that best "explains" the empirical phenomenon of the "law of small numbers."

The purpose of our note is to present an elementary and relatively simple

Received November 6, 1959.

[1] This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-43). This paper in whole or in part may be reproduced for any purpose of the United States Government.

proof of a bound of Kolmogorov's type. By using special features of the Poisson distribution, we are able to get the improved bound $3\sqrt[3]{\alpha}$ for D, and to accomplish this in a good deal simpler way than is required for the general result. We believe that our proof is suitable for presentation to an introductory class in probability theory.

2. The approximation theorems. Let X_i indicate success on the ith trial, so that $P(X_i = 1) = p_i$ and $P(X_i = 0) = 1 - p_i$. Our proofs will be based on the device of introducing random variables Y_i that have the Poisson distribution with $E(Y_i) = p_i$, and are such that $P(X_i = Y_i)$ is as large as possible. Specifically, we give to X_i and Y_i the joint distribution according to which

$$P(X_i = Y_i = 1) = p_i e^{-p_i}, \qquad P(X_i = 1, Y_i = 0) = p_i(1 - e^{-p_i}),$$

$$P(X_i = Y_i = 0) = e^{-p_i} - p_i(1 - e^{-p_i}),$$

and

$$P(X_i = 0, Y_i = y) = p_i^y e^{-p_i}/y! \qquad \text{for } y = 2, 3, \cdots.$$

We let the Y_i be independent of each other. (The construction is valid if $p_i \leqq 0.8$, insuring $P(X_i = Y_i = 0) \geqq 0$. For $p_i > 0.8$ the results below are trivially correct.)

From the familiar additive property of Poisson variables, we know that $T = \Sigma Y_i$ has exactly the Poisson distribution $\mathcal{P}(\Sigma p_i)$. Our objective is to show that $S = \Sigma X_i$ has nearly this distribution. Specifically, if we let

$$D = \sup_u | P(S \leqq u) - P(T \leqq u) |$$

denote the maximum absolute difference between the cumulatives of S and T, we want to find conditions under which D is small.

THEOREM 1. $D \leqq 2\Sigma p_i^2$.

Using the inequality $e^{-p_i} \geqq 1 - p_i$, it is easy to check that

$$P(X_i \neq Y_i) = 1 + p_i - (1 + 2p_i)e^{-p_i} \leqq 2p_i^2.$$

Therefore, by Boole's inequality, $P(S \neq T) \leqq \Sigma P(X_i \neq Y_i) \leqq 2\Sigma p_i^2$. But since $| P(S \leqq u) - P(T \leqq u) | \leqq P(S \neq T)$, the theorem follows.

In order to prove our next theorem, we shall need a uniform bound on the individual terms $p(k, \lambda) = e^{-\lambda}\lambda^k/k!$ of the Poisson distribution. It is well known that for large λ, the maximum term is of the order $\lambda^{-1/2}$, but we will give a specific upper bound.

LEMMA. *The maximum term of the distribution* $\mathcal{P}(\lambda)$ *is less than* $(1 + 1/12\lambda)/(2\pi\lambda)^{1/2}$.

PROOF. Suppose $k \leqq \lambda < k + 1$. The maximum term is then $e^{-\lambda}\lambda^k/k!$, as may be seen by looking at the ratio of successive terms. Since $(\lambda)^{1/2}e^{-\lambda}\lambda^k$ is maximized at $\lambda = k + \frac{1}{2}$, and since $1 + 1/12\lambda > 1 + 1/12(k + 1)$, it will suffice to show that

$$(2\pi)^{1/2}(k + \tfrac{1}{2})^{k+1/2}e^{-k-1/2} < k![1 + 1/12(k + 1)]$$

for $k = 0, 1, 2, \cdots$. This inequality may easily be checked by direct computation for $k = 0$, 1, and 2, and for $k \geqq 3$ by using the Stirling bound

$$k! > (2\pi)^{\frac{1}{2}} k^{k+\frac{1}{2}} e^{-k+(1/12k)-(1/360k^3)}.$$

Let us denote Σp_i by λ and Σp_i^2 by μ.

THEOREM 2. $D \leqq (3\mu/a^2) + (a + 1)(1 + 1/12\lambda)/(2\pi\lambda)^{\frac{1}{2}}$.

To prove this, we shall consider the random variables $Z_i = Y_i - X_i$.

$$E(Z_i) = 0,$$

while

$$\mathrm{Var}(Z_i) = E(Z_i^2) = p_i(1 - e^{-p_i}) + \sum_{k=2}^{\infty} k^2(p_i^k e^{-p_i})/k!$$

$$= p_i(1 - e^{-p_i}) + E(Y_i^2) - p_i e^{-p_i} = p_i^2 + 2p_i(1 - e^{-p_i}) \leqq 3p_i^2.$$

Let $\Sigma Z_i = U$. Then $E(U) = 0$ and $\mathrm{Var}(U) \leqq 3\mu$.

Let a be any positive number. If $T = S + U \leqq v - a$, then either $S \leqq v$ or $U \leqq -a$, so that $P(T \leqq v - a) \leqq P(S \leqq v) + P(U \leqq -a)$ and

$$P(T \leqq v) - P(S \leqq v) \leqq P(v - a \leqq T \leqq v) + P(U \leqq -a).$$

Similarly, if $S = T - U \leqq v$, then either $T \leqq v + a$ or $U \geqq a$, so that

$$P(S \leqq v) \leqq P(T \leqq v + a) + P(U \geqq a)$$

and $P(S \leqq v) - P(T \leqq v) \leqq P(v \leqq T \leqq v + a) + P(U \geqq a)$. Combining, we see that

$$D = \sup_v | P(S \leqq v) - P(T \leqq v) |$$

$$\leqq \sup_v P(v \leqq T \leqq v + a) + P(| U | \geqq a).$$

By the Chebycheff inequality, $P(| U | \geqq a) \leqq \mathrm{Var}\,(U)/a^2 \leqq 3\mu/a^2$. Using the lemma, we see that

$$\sup_v P(v \leqq T \leqq v + a) \leqq (a + 1)(1 + (1/12\lambda))/(2\pi\lambda)^{\frac{1}{2}},$$

since there are at most $a + 1$ Poisson terms in the interval from v to $v + a$. This completes the proof.

We now combine Theorems 1 and 2 to obtain our main result.

THEOREM 3. $D \leqq 3\sqrt[3]{\alpha}$.

We prove this by considering two cases. If $2\mu \leqq 3\sqrt[3]{\alpha}$, the theorem is an immediate consequence of Theorem 1. On the other hand, if $2\mu > 3\sqrt[3]{\alpha}$, we have by virtue of $\mu \leqq \alpha\lambda$ the inequality $\lambda > 3/2\alpha^{2/3} > 1$. Now suppose that $a \geqq 1$. Then

$$[(a + 1)(1 + (1/12\lambda))]/[(2\pi\lambda)^{\frac{1}{2}}] < a/\lambda^{\frac{1}{2}}$$

and

$$D \leqq (3\mu/a^2) + (a/\lambda^{\frac{1}{2}}).$$

This is minimized when $a = \sqrt[3]{6\mu(\lambda)^{\frac{1}{2}}} = a_0$. Since $\lambda > \mu/\alpha$ and $\mu > 3(\alpha)^{\frac{1}{2}}/2$, we see that $\mu(\lambda)^{\frac{1}{2}} > (\frac{3}{2})^{3/2}$ or $a_0 > (3^5/2)^{1/6} > 1$, so the restriction $a \geqq 1$ is satisfied, and the theorem is proved.

3. Remarks.

(i) We have presented our results as approximation theorems rather than as limit theorems. We believe it is better pedagogy to do so, since in the applications there will be definite values of n and the p_i, which are not "tending" to anything. However, if limit theorems are desired they follow at once. For example, Theorem 1 implies that $D \to 0$ as $\mu = \Sigma p_i^2 \to 0$, whereas Theorem 3 implies that $D \to 0$ as $\alpha = \max \{p_1, \cdots, p_n\} \to 0$.

(ii) Our Theorem 1 gives a simple and elementary proof of the standard textbook result that $\mathcal{B}(n, p) \to \mathcal{P}(\lambda)$ as $n \to \infty$, $p \to 0$, and $np \to \lambda$, since under these conditions $\Sigma p_i^2 = \lambda p \to 0$. Furthermore, Theorem 1 implies the more realistic theorem of von Mises, since if $\Sigma p_i = \lambda$ is fixed while $\alpha \to 0$, we must have $\Sigma p_i^2 \leqq \alpha \Sigma p_i = \alpha\lambda \to 0$.

(iii) As is customarily the case with bounds for the accuracy of approximations, our bound has only theoretical interest, being much too crude for practical usefulness. By pushing the method of proof, the constant factor 3 in the inequality $D \leqq 3\sqrt[3]{\alpha}$ can be reduced, but the result would still be of only theoretical value. It can be shown [3], using a much less simple argument, that $D \leqq 9\alpha$. While it is clearly a theoretical improvement to have a bound of order α rather than one of order $\sqrt[3]{\alpha}$, even the bound 9α is of limited applicational use. Fortunately, approximations are usually found in practice to be much better than the known bounds would indicate them to be.

(iv) The condition that $\alpha \to 0$ is sufficient but not necessary for $D \to 0$. It is easy to see that S will have approximately a Poisson distribution even if a few of the p_i are quite large, provided these values contribute only a small part of the total Σp_i.

REFERENCES

[1] William Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 2nd ed., New York, John Wiley and Sons, 1957.

[2] A. N. Kolmogorov, "Deux théorèmes asymptotiques uniformes pour sommes des variables aléatoires," *Teor. Veroyatnost. i Primenen.*, Vol. 1 (1956), pp. 426–436.

[3] Lucien Le Cam, "An approximation theorem for Poisson binomial distributions," to be published in the *Pacific J. Math.*

[4] R. v. Mises, "Über die Wahrscheinlichkeit seltener Ereignisse," *Z. Angew. Math. Mech.*, Vol. 1 (1921), pp. 121–124.

[5] Richard v. Mises, *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*, Wien, Deuticke, 1931.

CONFIDENCE BOUNDS CONNECTED WITH ANOVA AND MANOVA FOR BALANCED AND PARTIALLY BALANCED INCOMPLETE BLOCK DESIGNS[1]

BY V. P. BHAPKAR[2]

University of North Carolina

1. Introduction and summary. It is well known [3, 4] how, in the case of any general strongly testable [5] linear hypothesis for either ANOVA or MANOVA one can put simultaneous confidence bounds on a particular set of parametric functions, which might be regarded as measures of deviation from the "total" hypothesis and its various components. The parametric functions are such that, in each problem, one of these can be appropriately called the "total" and the rest "partials" of various orders. For each problem the "total" function, (i) in the univariate case, is related to, but not quite the same as, the noncentrality parameter of the usual F-test of the "total" hypothesis in ANOVA, and (ii) in the multivariate case, is the largest characteristic root of a certain parametric matrix which is related to, but not quite the same as, another parametric matrix whose nonzero characteristic roots occur as a set of noncentrality parameters in the power function for the test (no matter which of the standard tests we use) of the "total" hypothesis in MANOVA. The same remark applies to "partials" of various orders considered in the proper sense.

In this note, for both ANOVA and MANOVA, the hypothesis considered is that of equality of the treatment effects—vector equality in the case of MANOVA. Starting from such a hypothesis, explicit algebraic expressions are obtained for the total and partial parametric functions that go with the simultaneous confidence statements in the case of both ANOVA and MANOVA and for balanced and partially balanced designs. It is also indicated how to obtain, in a convenient form, the algebraic expression for the confidence bounds on each such parametric function, without a derivation of these expressions in an explicit form.

2. Notation and preliminaries.

(i) *Univariate case.* Let $\mathbf{x}$ denote a column-vector of n independent normal variables with a common variance σ^2 and the means given by

$$\mathcal{E}\mathbf{x} = \mathbf{A}_{n \times m}\boldsymbol{\theta}_{m \times 1}, \tag{1}$$

where $\mathbf{A}$ is a matrix of known constants and $\boldsymbol{\theta}$ is a vector of unknown parameters.

[1] This research was supported partly by the Office of Naval Research under Contract No. Nonr-855 (06) for research in probability and statistics at Chapel Hill and partly by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49 (638)-213. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Present address: Department of Statistics, University of Poona, Poona, 7, India.

The hypothesis

(2) $$\mathcal{H}_0: \quad \mathbf{B}_{s \times m}\theta = \mathbf{0} \qquad [\text{Rank} \quad \mathbf{B} = s]$$

is said to be strongly testable if Rank $(\mathbf{A}', \mathbf{B}')$ is equal to Rank $(\mathbf{A})$. If we write

(3) $$\mathbf{B}\theta = \phi_{s \times 1},$$

then the "total" parametric function, Δ, associated with $\mathcal{H}_0$ is

(4) $$\Delta = \phi' \mathbf{D}^{-1} \phi,$$

where $D\sigma^2$ is the variance-covariance matrix of the best unbiased linear estimates of ϕ. It may be observed that Δ/σ^2 is the noncentrality parameter of the F-test for $\mathcal{H}_0$. Confidence bounds on Δ, with a confidence coefficient greater than or equal to $(1 - \alpha)$, are then [3, 4] given by

(5) $$S_{H_0}^{\frac{1}{2}} - \left[\frac{s}{n-r} F_\alpha\right]^{\frac{1}{2}} S_E^{\frac{1}{2}} \leq \Delta^{\frac{1}{2}} \leq S_{H_0}^{\frac{1}{2}} + \left[\frac{s}{n-r} F_\alpha\right]^{\frac{1}{2}} S_E^{\frac{1}{2}},$$

where $r = \text{Rank } \mathbf{A}$, F_α is the $100\alpha\%$ significance point of F with d.f. s and $n - r$ respectively, S_{H_0} is the sum of squares due to $\mathcal{H}_0$ and S_E is the sum of squares due to error. We also have the simultaneous confidence statements

(6) $$S_{(a)H_0}^{\frac{1}{2}} - \left[\frac{s}{n-r} F_\alpha\right]^{\frac{1}{2}} S_E^{\frac{1}{2}} \leq \Delta_{(a)}^{\frac{1}{2}} \leq S_{(a)H_0}^{\frac{1}{2}} + \left[\frac{s}{n-r} F_\alpha\right]^{\frac{1}{2}} S_E^{\frac{1}{2}},$$

where $\Delta_{(a)} = \phi_{(a)}' \mathbf{D}_{(a)}^{-1} \phi_{(a)}$, $\phi_{(a)}$ is any subvector of ϕ, $\mathbf{D}_{(a)}$ is the corresponding submatrix of $\mathbf{D}$ and $S_{(a)H_0}$ is the corresponding sum of squares due to the partial hypothesis $\mathcal{H}_{(a)0}$: $\phi_{(a)} = \mathbf{0}$. (5) and (6) are implications of (13.2.21) on p. 90 in [3].

In the case of treatment-block designs, we have

(7) $$\mathcal{E}x_\alpha = t_i + b_j \qquad \begin{array}{l} i = 1, 2, \cdots, v, \\ j = 1, 2, \cdots, b, \end{array}$$

if the αth observation belongs to the ith treatment and jth block. The hypothesis of equality of treatment effects may be expressed as

(8) $$\mathcal{H}_0: \quad (\mathbf{I}_{v-1}, -\mathbf{J}_{v-1,1})\mathbf{t} = \mathbf{0},$$

where $\mathbf{t}' = (t_1, t_2, \cdots, t_v)$ and $\mathbf{J}_{r,s} = \{1\}_{r \times s}$. We shall write $\mathbf{J}_{r \times r}$ as $\mathbf{J}_r$. We assume that the design is connected. Let $n_{ij} = 1(0)$ if the ith treatment appears (does not appear) in the jth block. Then $\mathbf{N} = (n_{ij})_{v \times b}$ is the incidence-matrix of the design. Let r, k, $\mathbf{T}$ and $\mathbf{B}$ denote the number of replications of each treatment, the number of observations in each block, the vector of treatment totals and the vector of block-totals respectively. Then it is well-known [2] that the equations for $\mathbf{t}$ are

(9) $$\mathbf{C}\mathbf{t} = \mathbf{Q},$$

where $\mathbf{C} = r\mathbf{I} - (1/k)\mathbf{NN}'$ and $\mathbf{Q} = \mathbf{T} - (1/k)\mathbf{NB}$. Also

$$\text{Cov}(\mathbf{Q}) = \sigma^2\mathbf{C}. \tag{10}$$

Then, from (3) and (8), $\phi_i = t_i - t_v$, $i = 1, 2, \cdots, v-1$. We may express Δ in a symmetrical form by taking $\boldsymbol{\phi} = (\mathbf{I}_{v-1}, -\mathbf{J}_{v-1,1})\boldsymbol{\xi}$, where

$$\xi_i = t_i - (1/v)(t_1 + t_2 + \cdots + t_v), \qquad i = 1, 2, \cdots, v.$$

From (4)

$$\Delta = \boldsymbol{\xi}'(\mathbf{I}_{v-1}, -\mathbf{J}_{v-1,1})'\mathbf{D}^{-1}(\mathbf{I}_{v-1}, -\mathbf{J}_{v-1,1})\boldsymbol{\xi}. \tag{11}$$

(ii) *Multivariate case.* Let $\mathbf{X}$ denote a matrix of n independent p-dimensional normal variables with a common variance-covariance matrix $\boldsymbol{\Sigma}$, p being the number of characters observed on each individual, and let the means be given by

$$\mathcal{E}\mathbf{X}_{n\times p} = \mathbf{A}_{n\times m}\boldsymbol{\Theta}_{m\times p}, \tag{12}$$

where $\boldsymbol{\Theta}$ is a matrix of unknown parameters. Suppose that

$$\mathcal{H}_0: \quad \mathbf{B\Theta U}_{p\times u} = \mathbf{0} \qquad [\text{Rank } \mathbf{U} = u \leq p] \tag{13}$$

is the "strongly testable" hypothesis to be tested. If we write

$$\mathbf{B\Theta U} = \boldsymbol{\phi}_{s\times u}, \tag{14}$$

then the "total" parametric function, Δ, associated with $\mathcal{H}_0$ is [3, 4] given by

$$\Delta = C_{\max}[\boldsymbol{\phi}'\mathbf{D}^{-1}\boldsymbol{\phi}]. \tag{15}$$

It may be observed that the characteristic roots of $\boldsymbol{\phi}'\mathbf{D}^{-1}\boldsymbol{\phi}(\mathbf{U}'\boldsymbol{\Sigma}\mathbf{U})^{-1}$ are the noncentrality parameters in the power function of the test (no matter which of the standard tests we use) of the "total" hypothesis given by (13).

The confidence statement is [3] given by

$$\begin{aligned} C^{\frac12}_{\max}(\mathbf{S}_{H_0}) - \left[\frac{s}{n-r}C_\alpha\right]^{\frac12} C^{\frac12}_{\max}(\mathbf{S}_E) &\leq \Delta^{\frac12} \leq \\ C^{\frac12}_{\max}(\mathbf{S}_{H_0}) &+ \left[\frac{s}{n-r}C_\alpha\right]^{\frac12} C^{\frac12}_{\max}(\mathbf{S}_E), \end{aligned} \tag{16}$$

where $\mathbf{S}_{H_0}$ and $\mathbf{S}_E$ are the sum of products matrices due to the hypothesis and error respectively, and C_α is the $100\alpha\%$ significance point of the distribution of the largest characteristic root, with d.f. u, s, and $n - r$. In this case, we have simultaneous confidence statements, similar to (6), given by

$$\begin{aligned} C^{\frac12}_{\max}[\mathbf{S}_{(a)H_0}] - \left[\frac{s}{n-r}C_\alpha\right]^{\frac12} C^{\frac12}_{\max}(\mathbf{S}_E) &\leq \Delta^{\frac12}_{(a)} \\ &\leq C^{\frac12}_{\max}[\mathbf{S}_{(a)H_0}] + \left[\frac{s}{n-r}C_\alpha\right]^{\frac12} C^{\frac12}_{\max}(\mathbf{S}_E), \end{aligned} \tag{17}$$

where $\Delta_{(a)} = C_{\max}[\boldsymbol{\phi}'_{(a)}\mathbf{D}^{-1}_{(a)}\boldsymbol{\phi}_{(a)}]$, $\boldsymbol{\phi}_{(a)}$ being a submatrix of $\boldsymbol{\phi}$ obtained by choosing

some rows of ϕ. In addition, we have, by dropping some columns of ϕ, simultaneous confidence statements given by

$$(18)\quad \begin{aligned} C^{\frac{1}{2}}_{\max}[\mathbf{S}_{(b)H_0}] - \left[\frac{s}{n-r}C_\alpha\right]^{\frac{1}{2}} C^{\frac{1}{2}}_{\max}[\mathbf{S}_{(b)E}] &\leq \Delta^{\frac{1}{2}}_{(b)} \\ &\leq C^{\frac{1}{2}}_{\max}[\mathbf{S}_{(b)H_0}] + \left[\frac{s}{n-r}C_\alpha\right]^{\frac{1}{2}} C^{\frac{1}{2}}_{\max}[\mathbf{S}_{(b)E}], \end{aligned}$$

where $\Delta_{(b)} = C_{\max}[\phi'_{(b)}\mathbf{D}^{-1}\phi_{(b)}]$, $\phi_{(b)}$ being a submatrix of ϕ obtained by choosing some columns of ϕ, and $\mathbf{S}_{(b)H_0}$ and $\mathbf{S}_{(b)E}$ are the corresponding submatrices of $\mathbf{S}_{H_0}$ and $\mathbf{S}_E$. (16), (17) and (18) are implications of (14.6.3) on p. 101 in [.3]

In the case of treatment-block designs, we have

$$(19)\qquad \mathcal{E}x_\alpha^{(k)} = t_i^{(k)} + b_j^{(k)}, \qquad \begin{aligned} i &= 1, 2, \cdots, v, \\ j &= 1, 2, \cdots, b, \\ k &= 1, 2, \cdots, p, \end{aligned}$$

where $x_\alpha^{(k)}$ denotes the kth character measured on the αth experimental unit or individual that turns up for the ith treatment and the jth block; and $t_i^{(k)}$, $b_j^{(k)}$ stand respectively for the contributions to the expectation of the kth variate made by the ith treatment and the jth block.

From (1) and (12) we have the same "structure matrix", $\mathbf{A}$, in the multivariate situation as in the univariate case. This "structure matrix" depends on the design as well as on what the experimental statisticians have called the model, e.g., (7) and (19).

In this set-up, so far as the hypothesis (13) is concerned, we shall take $\mathbf{U} = \mathbf{I}$ for simplicity.

3. Balanced incomplete block designs.

(i) *Univariate case.* Here

$$\mathbf{C} = r\mathbf{I}_v - \frac{1}{k}[(r-\lambda)\mathbf{I}_v + \lambda\mathbf{J}_v] = \frac{\lambda v}{k}\mathbf{I}_v - \frac{\lambda}{k}\mathbf{J}_v.$$

Imposing the usual condition, $\mathbf{J}_v\mathbf{t} = \mathbf{0}$, to get unique solutions, we have $\mathbf{t} = (k/\lambda v)\mathbf{Q}$. Therefore, $\phi = k/\lambda v(\mathbf{I}_{v-1}, -\mathbf{J}_{v-1,1})\mathbf{Q}$, and hence

$$(20)\qquad \mathbf{D} = \frac{k^2}{\lambda^2 v^2}(\mathbf{I}_{v-1}, -\mathbf{J}_{v-1,1})\mathbf{C}(\mathbf{I}_{v-1}, -J_{v-1,1})' = \frac{k}{\lambda v}(\mathbf{I}_{v-1} + \mathbf{J}_{v-1}),$$

whence

$$(21)\qquad \mathbf{D}^{-1} = \frac{\lambda v}{k}(\mathbf{I}_{v-1} - (1/v)\mathbf{J}_{v-1}).$$

Thus, using (11) and the relation $\mathbf{J}_v\xi = \mathbf{0}$,

$$\Delta = \frac{\lambda v}{k} \phi' \left(\mathbf{I}_{v-1} - \frac{1}{v} \mathbf{J}_{v-1} \right) \phi$$

$$= \frac{\lambda v}{k} \xi' \left(\mathbf{I}_v - \frac{1}{v} \mathbf{J}_v \right) \xi$$

$$= \frac{\lambda v}{k} \xi' \xi \tag{22}$$

$$= \frac{\lambda v}{k} \sum_{i=1}^{v} \xi_i^2.$$

Then we can have a confidence statement of the form of (5) with $n = bk$, $r = b + v - 1$, $s = v - 1$ and Δ given by (22).

For the "partial" statements (6), if $\phi'_{(a)} = (\phi_{i_1}, \phi_{i_2}, \cdots, \phi_{i_t})$, $(t < v - 1)$ then, from (20),

$$\mathbf{D}_{(a)} = \frac{k}{\lambda v} (\mathbf{I}_t + \mathbf{J}_t). \tag{23}$$

Hence

$$\mathbf{D}_{(a)}^{-1} = \frac{\lambda v}{k} \left(\mathbf{I}_t - \frac{1}{t+1} \mathbf{J}_t \right) \tag{24}$$

and

$$\Delta_{(a)} = \frac{\lambda v}{k} \phi'_{(a)} \left(\mathbf{I}_t - \frac{1}{t+1} \mathbf{J}_t \right) \phi_{(a)}.$$

For a symmetrical expression, we take $\phi_{(a)} = (\mathbf{I}_t, -\mathbf{J}_{t,1}) \xi_{(a)}$, where

$$\xi_{i_j(a)} = t_{i_j} - (1/(t+1))[t_{i_1} + t_{i_2} + \cdots + t_{i_t} + t_v]$$

so that, using $\mathbf{J}_{t+1} \xi_{(a)} = 0$,

$$\Delta_{(a)} = \frac{\lambda v}{k} \xi'_{(a)} \left(\mathbf{I}_{t+1} - \frac{1}{t+1} \mathbf{J}_{t+1} \right) \xi_{(a)}$$

$$= \frac{\lambda v}{k} \xi'_{(a)} \xi_{(a)} \tag{25}$$

$$= \frac{\lambda v}{k} \left[\sum_{j=1}^{t} \xi_{i_j(a)}^2 + \xi_{v(a)}^2 \right].$$

(ii) *Multivariate case.* We have the confidence bounds of the form of (16) with $n = bk$, $r = b + v - 1$, $s = v - 1$ and

$$\Delta = C_{\max} \left[\phi' \frac{\lambda v}{k} \left(\mathbf{I}_{v-1} - \frac{1}{v} \mathbf{J}_{v-1} \right) \phi \right].$$

Here again we may write $\phi = (\mathbf{I}_{v-1}, -\mathbf{J}_{v-1,1}) \xi$, where $\xi = (\xi^{(1)}, \cdots, \xi^{(p)})$ and

$\xi_i^{(j)} = t_i^{(j)} - v^{-1}\sum_{i=1}^{v} t_i^{(j)}$. Then $\Delta = (\lambda v/k)C_{\max}(\xi'\xi)$. We have, from (24), one set of "partial" statements of the form of (17) with

$$\Delta_{(a)} = \frac{\lambda v}{k} C_{\max}\left[\phi'_{(a)}\left(\mathbf{I}_t - \frac{1}{t+1}\mathbf{J}_t\right)\phi_{(a)}\right],$$

or, from (25), $\Delta_{(a)} = (\lambda v/k)C_{\max}[\xi'_{(a)}\xi_{(a)}]$, where $\xi_{(a)} = (\xi_{(a)}^{(1)}, \cdots, \xi_{(a)}^{(p)})$.

Similarly, we have, from (21), another set of "partial" statements of the form of (18) with

$$\Delta_{(b)} = C_{\max}\left[\phi'_{(b)}\frac{\lambda v}{k}\left(\mathbf{I}_{v-1} - \frac{1}{v}\mathbf{J}_{v-1}\right)\phi_{(b)}\right] = \frac{\lambda v}{k}C_{\max}(\xi'_{(b)}\xi_{(b)}).$$

4. Partially balanced incomplete block designs.

(i) *Univariate case.* Consider a PBIBD with m associate classes and association matrices $\mathbf{B}_i$ $(i = 0, 1, \cdots, m)$. Then it is well known [1] that $\mathbf{C} = \sum_{k=0}^{m}\alpha_k\mathbf{B}_k$, where $\mathbf{B}_0 = \mathbf{I}_v$, $\alpha_0 = r(k-1)/k$, $\alpha_i = -\lambda_i/k$, $i = 1, \cdots, m$; and, imposing the condition $\mathbf{J}_{1,v}\mathbf{t} = 0$ on (9), we have

$$\mathbf{t} = \left(\sum_{k=0}^{m} e_k\mathbf{B}_k\right)\mathbf{Q} = \mathbf{E}\mathbf{Q}, \quad \text{say}.$$

It is well known that, when the design is connected, Rank $\mathbf{C} = v - 1$, so that the condition $\mathbf{J}_{1,v}\mathbf{t} = 0$ is sufficient to give unique solutions. Further

$$\mathbf{J}_{1,v}\mathbf{C} = 0 \quad \text{and} \quad \mathbf{J}_{1,v}\mathbf{Q} = 0. \tag{26}$$

Let

$$\underset{v\times v}{\mathbf{C}} = \begin{pmatrix}\mathbf{C}_1\\ \mathbf{c}'\end{pmatrix}\begin{matrix}v-1\\ 1\end{matrix} = (\mathbf{C}'_1, \mathbf{c}) = \begin{pmatrix}\mathbf{C}_{11} & \mathbf{d}\\ \mathbf{d}' & c_0\end{pmatrix},$$

and

$$\mathbf{E} = \begin{pmatrix}\mathbf{E}_1\\ \mathbf{e}'\end{pmatrix}\begin{matrix}v-1\\ 1\end{matrix} = (\mathbf{E}'_1, \mathbf{e}) = \begin{pmatrix}\mathbf{E}_{11} & \mathbf{f}\\ \mathbf{f}' & e_0\end{pmatrix}.$$

Then $\begin{pmatrix}\mathbf{C}_1\\ \mathbf{J}_{1,v}\end{pmatrix}\mathbf{t} = \begin{pmatrix}\mathbf{Q}_1\\ 0\end{pmatrix}$, where $\mathbf{Q}' = (\mathbf{Q}'_1, Q_v)$. Hence

$$\mathbf{t} = \mathbf{E}\mathbf{Q} = (\mathbf{E}'_1, \mathbf{e})\begin{pmatrix}\mathbf{Q}_1\\ Q_v\end{pmatrix} = \mathbf{E}'_1\mathbf{Q}_1 + \mathbf{e}Q_v.$$

Therefore, in view of (26),

$$\mathbf{t} = \mathbf{E}'_1\mathbf{Q}_1 - \mathbf{e}\mathbf{J}_{1,v-1}\mathbf{Q}_1 = (\mathbf{E}'_1 - \mathbf{e}\mathbf{J}_{1,v-1})\mathbf{Q}_1 = (\mathbf{E}'_1 - \mathbf{e}\mathbf{J}_{1,v-1} \mid \mathbf{x})\begin{pmatrix}\mathbf{Q}_1\\ 0\end{pmatrix}.$$

Hence

$$\begin{pmatrix}\mathbf{C}_1\\ \mathbf{J}_{1,v}\end{pmatrix}^{-1} = (\mathbf{E}'_1 - \mathbf{e}\mathbf{J}_{1,v-1} \mid \mathbf{x}), \tag{27}$$

so that,

$$(\mathbf{E}_1' - \mathbf{e}\mathbf{J}_{1,v-1} \,\vdots\, \mathbf{x}) \begin{pmatrix} \mathbf{C}_1 \\ \mathbf{J}_{1,v} \end{pmatrix} = \mathbf{I}_v .$$

Thus $\mathbf{E}_1'\mathbf{C}_1 - \mathbf{e}\mathbf{J}_{1,v-1}\mathbf{C}_1 + \mathbf{x}\mathbf{J}_{1,v} = \mathbf{I}_v$. Hence, in view of (26),

$$\mathbf{E}_1'\mathbf{C}_1 + \mathbf{e}\mathbf{C}' = \mathbf{I}_v - \mathbf{x}\mathbf{J}_{1,v} ,$$

that is,

$$\mathbf{E}\mathbf{C} = \mathbf{I}_v - \mathbf{x}\mathbf{J}_{1,v} . \tag{28}$$

Also, from (27),

$$\begin{pmatrix} \mathbf{C}_1 \\ \mathbf{J}_{1,v} \end{pmatrix} (\mathbf{E}_1' - \mathbf{e}\mathbf{J}_{1,v-1} \,\vdots\, \mathbf{x}) = \mathbf{I}_v .$$

Hence $\mathbf{C}_1\mathbf{x} = \mathbf{0}$. But $\mathbf{C}_1\mathbf{J}_{v,1} = \mathbf{0}$ and Rank $\mathbf{C}_1 = v - 1$. Therefore, $\mathbf{x} = x\mathbf{J}_{v,1}$. Furthermore, $\mathbf{J}_{1,v}\mathbf{x} = 1$, whence $x\mathbf{J}_{1,v}\mathbf{J}_{v,1} = 1$, that is, $x = v^{-1}$. (28) thus reduces to

$$\mathbf{E}\mathbf{C} = \mathbf{I}_v - (1/v)\mathbf{J}_v . \tag{29}$$

Now $\hat{\boldsymbol{\phi}} = (\mathbf{I}_{v-1}, -\mathbf{J}_{v-1,1})\mathbf{t} = (\mathbf{I}_{v-1}, -\mathbf{J}_{v-1,1})\mathbf{E}\mathbf{Q}$, so that,

$$\mathbf{D} = (\mathbf{I}_{v-1}, -\mathbf{J}_{v-1,1})\mathbf{E}\mathbf{C}\mathbf{E}\begin{pmatrix} \mathbf{I}_{v-1} \\ -\mathbf{J}_{1,v-1} \end{pmatrix}.$$

Therefore, from (29) and (27),

$$\begin{aligned} \mathbf{D} &= (\mathbf{I}_{v-1}, -\mathbf{J}_{v-1,1})\left(\mathbf{I}_v - \frac{1}{v}\mathbf{J}_v\right)(\mathbf{E}_1' - \mathbf{e}\mathbf{J}_{1,v-1}) \\ &= (\mathbf{I}_{v-1}, -\mathbf{J}_{v-1,1})(\mathbf{E}_1' - \mathbf{e}\mathbf{J}_{1,v-1}) \\ &= \mathbf{E}_{11} - \mathbf{f}\mathbf{J}_{1,v-1} - \mathbf{J}_{v-1,1}\mathbf{f}' + e_0\mathbf{J}_{v-1} . \end{aligned} \tag{30}$$

Furthermore, premultiplying both sides of the equation,

$$\left[\begin{array}{c:c} \mathbf{E}_{11} - \mathbf{f}\mathbf{J}_{1,v-1} & \frac{1}{v}\mathbf{J}_{v-1,1} \\ \mathbf{f}' - e_0\mathbf{J}_{1,v-1} & \frac{1}{v} \end{array}\right] \begin{bmatrix} \mathbf{C}_{11} & \mathbf{d} \\ \mathbf{J}_{1,v-1} & 1 \end{bmatrix} = \mathbf{I}_v ,$$

by $(\mathbf{I}_{v-1}, -\mathbf{J}_{v-1,1})$, we have $\mathbf{D}\mathbf{C}_{11} = \mathbf{I}_{v-1}$ and, therefore,

$$\mathbf{D}^{-1} = \mathbf{C}_{11} . \tag{31}$$

Hence, from (11),

$$\Delta = \boldsymbol{\phi}'C_{11}\boldsymbol{\phi} = \boldsymbol{\xi}'\mathbf{C}\boldsymbol{\xi}. \tag{32}$$

Here, we may note that $c_{ii} = \alpha_0 = r(k-1)/k$ and $c_{ij} = \alpha_l = -\lambda_l/k$ if ith and jth treatments are lth associates. Then we can have a confidence statement of the form of (5) with $n = bk$, $r = b + v - 1$, $s = v - 1$ and Δ given by (32).

The "partial" statements of the form of (6), however, cannot be made in a compact form, unless we know the association scheme. If we have $\boldsymbol{\phi}'_{(a)} = (\phi_1, \cdots, \phi_t)$, then $\mathbf{D}_{(a)}^{-1} = \mathbf{X} - \mathbf{Y}\mathbf{W}^{-1}\mathbf{Z}$, where

$$\mathbf{C}_{11} = \begin{array}{cc} \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Z} & \mathbf{W} \end{bmatrix} & \begin{array}{c} t \\ v-t-1 \end{array} \\ \begin{array}{cc} t & v-t-1 \end{array} & \end{array}$$

and thus $\Delta_{(a)} = \boldsymbol{\phi}'_{(a)}[\mathbf{X} - \mathbf{Y}\mathbf{W}^{-1}\mathbf{Z}]\boldsymbol{\phi}_{(a)}$.

(ii) *Multivariate case.* We have the confidence bounds of the form of (16) with $n = bk$, $r = b + v - 1$, $s = v - 1$ and

$$\Delta = C_{\max}[\boldsymbol{\phi}'\mathbf{C}_{11}\boldsymbol{\phi}] = C_{\max}[\boldsymbol{\xi}'\mathbf{C}\boldsymbol{\xi}].$$

We have, as before, one set of "partial" statements of the form of (17) with

$$\Delta_{(a)} = C_{\max}[\boldsymbol{\phi}'_{(a)}(\mathbf{X} - \mathbf{Y}\mathbf{W}^{-1}\mathbf{Z})\boldsymbol{\phi}_{(a)}].$$

The other set of "partial" statements is of the form of (18) with

$$\Delta_{(b)} = C_{\max}[\boldsymbol{\phi}'_{(b)}\mathbf{C}_{11}\boldsymbol{\phi}_{(b)}] = C_{\max}[\boldsymbol{\xi}'_{(b)}\mathbf{C}\boldsymbol{\xi}_{(b)}].$$

5. General "connected" incomplete block designs. It is well known [2] that, in general, $\mathbf{Ct} = \mathbf{Q}$, which, on imposing the condition $\mathbf{J}_{1,v}\mathbf{t} = 0$, yields $\mathbf{t} = \mathbf{EQ}$. Then, arguing as before, from (26) to (32), we have

$$\Delta = \boldsymbol{\phi}'\mathbf{C}_{11}\boldsymbol{\phi} = \boldsymbol{\xi}'\mathbf{C}\boldsymbol{\xi}. \tag{33}$$

Then we can have a confidence statement of the form of (5) with

$$n = \sum_{i=1}^{v} r_i = \sum_{j=1}^{b} k_j, \qquad r = b + v - 1, \quad s = v - 1$$

and Δ given by (33). We can have "partial" statements and confidence bounds in the multivariate situation analogous to those for PBIBD.

6. Acknowledgment. I am indebted to Professor S. N. Roy for suggesting this problem and for suggesting improvements.

REFERENCES

[1] R. C. Bose and Dale M. Mesner, "On linear associative algebras corresponding to association schemes of partially balanced designs," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 21–38.

[2] R. C. Bose and T. Shimamoto, "Classification and analysis of partially balanced incomplete block designs with two associate classes," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 151–184.

[3] S. N. Roy, *Some Aspects of Multivariate Analysis*, John Wiley and Sons, New York, 1957.

[4] S. N. Roy and J. Roy, "Analysis of variance with univariate or multivariate, fixed or mixed classical models," Institute of Statistics, University of North Carolina, Mimeograph Series No. 208, 1958.

[5] S. N. Roy and J. Roy, "On testability in normal ANOVA and MANOVA with all fixed effects," Jubilee No., *Bull. Calcutta Math. Soc.*, 1959.

A CLASS OF FACTORIAL DESIGNS WITH UNEQUAL CELL-FREQUENCIES

By Gideon Schwarz[1]

Columbia University

1. Summary. A class of multifactorial designs are defined and analyzed. The designs considered have each a total number of observations that can not be divided equally among the cells of the designs; however, by distributing the observations in a way that is in a certain sense symmetrical, the equations that determine the least squares estimates of the linear parameters become explicitly solvable.

The case of two non-interacting factors with arbitrary numbers of levels is treated first. In the n-factor case we have to restrict ourselves to factors having equal numbers of levels. After defining the designs, the estimates are computed. Some general discussions of the symmetries and algebraic properties involved conclude the paper.

2. Introduction. The first case to be considered is that of two non-interacting factors, with I and J levels respectively. For each pair i, j of levels the measured magnitude has an expected value η_{ij}. We assume that the η_{ij} can be expressed in terms of $I + J + 1$ parameters $\{\mu, \alpha_i, \beta_j\}$ by the equations

$$\eta_{ij} = \mu + \alpha_i + \beta_j, \qquad \alpha_{.} = \beta_{.} = 0. \tag{1}$$

The dot indicates as usual summation over the range of the index it replaces.

Denoting by y_{ijk} the kth measurement in the cell in which the factors A and B are applied at levels i and j respectively, we assume the y_{ijk} to be normal independent random variables with means η_{ij} and common variance σ^2.

The experimenter is free to choose the number n_{ij} of observations in each cell. The choice of the matrix n_{ij} may be influenced by three requirements; first, the cost of experimentation makes an unnecessarily large number of observations undesirable; second, for a given number $n_{..}$ of observations, different ways of dividing this number among the different cells will result in different patterns of information about the parameters, and unless specific conditions about some of the levels are added, the design will be the closer to optimal the more evenly the number $n_{..}$ is distributed among the cells; and last, it is impossible to write simple explicit formulas for the least-squares-estimates that hold for general n_{ij}, while for some classes of n_{ij}-matrices, such formulae can be found.

Considering the two last requirements only, we are led to a well known class of designs, namely those in which all the n_{ij} are equal, say to n.

Received July 17, 1959; revised April 19, 1960.

[1] This research was supported by the Office of Naval Research under Contract Number Nonr-266(59), Project Number 042-205. Reproduction in whole or in part is permitted for any purpose of the United States Government.

As we have $n_{..} = nIJ$ for this class, we cannot regulate the total number of experiments except in jumps of IJ; in many cases this may lead to a violation of the first requirement. Consider for example a case in which one observation per cell would suffice for estimation of the parameters μ, α_i, β_j, while for the estimation of σ^2, we would want a few additional observations in some of the cells. Within the class of constant n_{ij}, this can be achieved only by doubling the total number of experiments.

There have been various attempts of considering special designs with unequal frequencies (Cf. References). Among the special cases treated by Daniel [2] and, in private communications with Daniel, by A. Birnbaum and Scheffé, there were designs with some symmetry properties. It was Birnbaum's suggestion to look for a more general class of designs that led to the results described in this paper.

3. Definition of S and Calculation of the Estimates. Let us proceed now to define the class **S**. We start out with d by d unit matrix, d being any common divisor of I and J, and change it into an I by J matrix by replacing each of its "one" entries by a I/d by J/d matrix of ones, and each of its zeros by a similar matrix of zeros. This way we define a matrix

$$\left(\begin{array}{ccc|ccc} 1 & 1 & \cdots & & & \\ 1 & \cdot & & & & \\ 1 & & \cdot & & 0 & \cdots \\ \vdots & & & & & \\ \hline & & & 1 & 1 & \cdots \\ & & & 1 & & \\ & & 0 & 1 & & \\ & & & \vdots & & \\ \hline & & \vdots & 0 & & \ddots \end{array}\right)$$

Denoting this matrix by $A_{I,J,d}$, or for short $A_{I,J}$, we can now define **S** as the class of all designs with matrices (n_{ij}) that can be written either in the form $(n) + A_{I,J}$ or $(n) - A_{I,J}$ for some positive integer n, and d, a divisor of n, where (n) denotes the I by J matrix having all entries equal to n. We claim, (a) the number $n_{..}$ runs in the class S over all integers of the form

$$IJ(n \pm d^{-1});$$

and (b) there is a simple explicit formula for the least-squares estimates that holds for all the designs in **S**.

(a) becomes evident if we observe that $A_{I,J}$ has IJ/d non-zero entries, and we shall prove (b) by arriving at the formulae, first for the minus sign and then in general.

The least squares estimates of the row effects can be obtained from the numbers

$$a_i = y_{i..}/n_{i.} - y_{...}/n_{..}, \tag{2}$$

which span uniquely the estimation space restricted by the side conditions. Each a_i is a unique linear combination of the least squares estimates $\hat{\alpha}_i$, $\hat{\beta}_i$ given by

$$a_i = \hat{\alpha}_i + [1/J(n + d^{-1})] \sum{}^* \hat{\beta}_j, \tag{3}$$

where $\sum^*$ denotes summation over the cells with $n + 1$ observations only. Using vector notation $a = (a_1, \cdots, a_I)$, etc.,

$$a = \hat{\alpha} + [d/J(nd + 1)] A_{IJ}\hat{\beta}, \tag{4}$$

and, by interchanging rows and columns,

$$b = \hat{\beta} + [d/I(nd + 1)] A_{JI}\hat{\alpha}. \tag{5}$$

To eliminate $\hat{\beta}$ from equations (4) and (5), we subtract from (4) a suitable multiple of (5). Using the equation

$$A_{IJ}A_{JK} = (J/d)A_{IK} \tag{6}$$

which follows easily from the definition of A_{IJ}, we arrive at

$$a - [d/J(nd + 1)]A_{IJ}b = \hat{\alpha} - [d/I(nd + 1)^2]A_{II}\hat{\alpha}. \tag{7}$$

In order to solve this equation, we have to invert a matrix which can be written, if we denote the unit matrix by U_{II}, as $U_{II} - [d/I(nd + 1)^2]A_{II}$.

We can find the required inverse by finding the value of t that makes the product

$$(U_{II} - [d/I(nd + 1)^2]A_{II})\ (U_{II} + tA_{II}) \tag{8}$$

equal to the unit matrix. Reducing the A_{II}^2 term by applying (6) we obtain $t = 1/In(nd + 2)$. Having found the inverse we can now solve equation (7). Denoting by R and C vectors of row and column-sums, respectively, and by S a vector with I components, all equal to the grand total $y_{\dots}$, we have

$$\begin{aligned} \hat{\alpha} = {} & [d/J(nd + 1)]R + [d/IJn(nd + 1)(nd + 2)]A_{II}R \\ & - [d/IJn(nd + 2)]A_{IJ}C - [d/IJ(nd + 2)]S. \end{aligned} \tag{9}$$

The corresponding formula for $\hat{\beta}$ is easily obtained by interchanging R and C, as well as I and J. The estimate of μ is obviously equal to $(d/IJ(nd - 1))y\cdots$, the mean of all observations. The change in the formulae for the case $(n_{ij}) = (n) - A_{I,J}$, will consist of changing the signs of d, and of all the matrices. Merging both cases into one, and denoting by S also a vector with J components all equal to $y\cdots$, we have finally

$$\begin{aligned} \hat{\alpha} = {} & [d/J(nd \pm 1)]R + [d/IJn(nd \pm 1)(nd \pm 2)A_{II}R \\ & \mp [d/IJn(nd \pm 2)]A_{IJ}C - [d/IJ(nd \pm 2)]S, \end{aligned} \tag{10}$$

$$\begin{aligned} \hat{\beta} = {} & [d/I(nd \pm 1)]C + [d/JIn(nd \pm 1)(nd \pm 2)]A_{JJ}C \\ & \mp [d/JIn(nd \pm 2)]A_{JI}R - [d/JI(nd \pm 2)]S, \end{aligned} \tag{11}$$

$$\hat{\mu} = [d/IJ(nd \pm 1)]S. \tag{12}$$

As a final remark, let us note that our definitions and formulae are valid as long as n is at least 1, and d is at least 2, with the exception of the case $n = 1$, $d = 2$, in which $nd - 2$ equals zero, and the $n - d^{-1}$ replicate is not sufficient for estimation of the parameters. On the other hand, as for $d \geqq 3$, $n = 1$ the formulae remain meaningful also for the lower sign, certain designs with some empty cells are included in the class considered here.

Most of what has been done in the preceeding section admits a rather straightforward generalization to the case of q factors acting additively, that is, with no interactions of any order. The only step that is not generalized so easily is the reduction of equations (4) and (5), each involving both row-effect estimates and column effect estimates, to equation (7), which isolates the row effects. In order to make possible an explicit solution to the analogous problem in the case of many factors, we have to restrict our considerations to designs having an equal number of levels for every factor. Denoting the effect of the hth factor at its ith level by $\alpha_{i(h)}$, we define our model by the equations

$$y_{i,k} = \mu + \sum_h \alpha_{i(h)} + \epsilon_{i,k} \tag{13}$$

where i denotes the vector $(i(1), \cdots i(q))$, and $k = 1, 2, \cdots, n_{i,j}$, with the error terms distributed as usual. About the parameters we assume $\sum_i \alpha_{i(h)} = 0$, $h = 1, 2, \cdots q$.

In order to determine the number of observations in each cell, we choose a divisor d of the number of levels I, and construct a q-dimensional hyper-cube of side-length d. Putting d ones at the grid points along the q-space-diagonal of the hyper-cube, and zeros at the other grid points, we obtain the q-dimensional analogue of the d by d unit matrix. Replacing each $(q - 1)$ dimensional layer by I/d identical layers, an array of I^q points is obtained, I^q/d^{q-1} of which carry units. If we start out with an I^q-design having n observations in each cell, and add ± 1 observation to each cell that corresponds to a unit in the array, an $n \pm 1/d^{q-1}$ duplicate will be obtained.

Defining the numbers $a_i(h)$ as the average of the observations in the layer determined by a given level of a given factor, minus the average of all observations, we get a system of vector equations;

$$\begin{aligned} a(1) &= \hat{\alpha}(1) + gA_{II}\hat{\alpha}(2) + gA_{II}\hat{\alpha}(3) + \cdots + gA_{II}\hat{\alpha}(q) \\ a(2) &= gA_{II}\hat{\alpha}(1) + \hat{\alpha}(2) + gA_{II}\hat{\alpha}(3) + \cdots + gA_{II}\hat{\alpha}(q) \\ &\vdots \\ a(q) &= gA_{II}\hat{\alpha}(1) + \cdots \qquad \cdots + gA_{II}\hat{\alpha}(q-1) + \hat{\alpha}(q) \end{aligned} \tag{14}$$

where $g = [d^{q-1}/I^{q-1}(nd^{q-1} \pm 1)](I/d)^{q-2} = d/I(nd^{q-1} \pm 1)$. The first factor in g is the reciprocal of the number of observations per $(n - 1)$ − dimensional partial design. The second factor is the number of higher populated cells that two levels of different factors have in common.

For the solution of (14) inversion of a q by q matrix having I by I matrices as elements (a q by q by I by I tensor of the fourth degree) is required. We start

by considering the matrix G_{qq} obtained by replacing A_{II} in the tensor of (14) by a scalar variable x. Putting for its inverse

$$G_{qq}^{-1} = \begin{pmatrix} z & y & y & \cdots & y \\ y & z & y & \cdots & y \\ \vdots & \vdots & \vdots & & \vdots \\ y & y & \cdots & y & z \end{pmatrix}$$

we find

$$y = gx/[(q-1)g^2x^2 - (q-2)gx - 1],$$
$$z = -(q-2)gx - 1/[(q-1)g^2x^2 - (q-2)gx - 1].$$

We can write this result in a form that does not involve any fractions

$$(15) \quad \begin{aligned} &G_{qq}[((q-1)g^2x^2 - (q-2)gx - 1)G_{qq}^{-1}] \\ &\qquad = ((q-1)g^2x^2 - (q-2)gx - 1)U_{qq}, \end{aligned}$$

where the expression in the square brackets equals a q by q matrix with $-(q-2)gx - 1$ along the diagonal, and gx in the other places. Having disposed of fractions, we can now substitute A_{II} for x. Carrying out the substitution in equation (15), G_{qq} becomes the tensor of (14), and the expression in the square brackets becomes a tensor having $-(q-2)gA_{II} - U_{II}$ along its diagonal, and gA_{II} in the other places.

Applying all this to (14), we arrive at the reduced equations,

$$(16) \quad \begin{aligned} &[(q-1)g^2A_{II}^2 - (q-2)gA_{II} - U_{II}]\,\hat{a}(1) \\ &\qquad = -[(q-2)gA_{II} - U_{II}]a(1) + gA_{II}[a(2) + \cdots] \end{aligned}$$

We can now proceed as in the two-factor case, and get, putting N for nd^{q-1},

$$(17) \quad \begin{aligned} \hat{a}(1) = {} & d^{q-1}/[I^{q-1}(N \pm 1)]S(1) \\ & + [(q-1)d^q/I^qN(N \pm 1)(N \pm q)]A_{II}S(1) \\ & \mp d^q/[I^qN(N-q)A_{II}[S(2) + \cdots S(q)] - d^{q-1}/I^q(N \pm q)S. \end{aligned}$$

As in the 2-factor case, the lower signs serve for the $n - 1/d^{q-1}$ duplicate. As for $q \geqq 3$ we have $N = nd^{q-1} > q$, the denominators never vanish except in the case mentioned before when $q = 2$ and $d = 2$, and the formulae are valid unrestrictedly. By permuting the factors, estimates for the other factors can be easily obtained.

4. General Symmetrical Designs. In this section we shall examine closer the symmetry properties that the designs treated in this paper have in common.

The various symmetry properties of the designs having equal numbers of observations in all cells are implied by the invariance of these designs under all permutations of the levels of any factors; furthermore, those designs, the "full multiple replicates", are the only ones left invariant by all permutations. Cer-

tainly, invariance under all permutations assures us of equal treatment of all levels of each factor. It implies an even stronger property: different ordered pairs of levels will enter the design similarly, as will any different ordered n-tuples. This additional property is certainly welcome. Some of the questions the designed experiment might be called upon to answer do involve pairs or other sets of levels, and it would be natural to expect symmetric treatment of these questions as well. However, we know that we have to give up some requirements if we want to include fractional replicates, and it is this "symmetry of subsets" that we choose to sacrifice.

Let us examine the freedom gained by requiring symmetry with regard to single levels only, by looking at the two-factor case. In this case, the design is determined by a matrix having the cell frequencies as its entries. Applying single-level symmetry to the row factor, we find that the rows of the matrix have to be equal to each other; however, as the order of entries in a row is determined by the order of levels of the column factor, the order in a row is immaterial, and the word "equal" should be read "differing only by a permutation of their elements". Similar "equality" is implied for the columns of the matrix. Any unit matrix can now serve as an example of a matrix having the required properties and yet not belonging to the full replicate designs.

Returning to the multifactorial designs, we arrive at the following formulation of our symmetry requirements:

Definition: A design is called "symmetrical with regard to single levels", or from here on, for short, "symmetrical", if the two partial designs resulting from fixing any one of the factors at two different levels, can be transformed one into the other by permuting the levels of the other factors.

We now restrict our class of designs even further, by introducing a restriction that is not motivated solely by considerations of symmetry. The designs we shall consider will all have only two different numbers of observations per cell occurring in their cells, furthermore, those two numbers will differ from each other by one. We justify this restriction by the following "optimality argument": the definition of a symmetrical design implies that the different cell frequencies appearing in partial designs belonging to different levels of the same factor, will be the same, possibly differently arranged. If there were two cells in the design whose numbers of observations differed by more than one, we would find two such cells in every subdesign and a new design could be defined by decreasing by one the number of observations in the higher populated cell and increasing it in the other. The resulting design would still be symmetrical and have the same total number of observations as the original design. As whenever the given total number of observations makes equally populated cells possible, the fully replicated design is in some sense optimal, we can interpret the above restriction as an attempt to avoid unnecessary deviations from optimality.

Having narrowed down the class of designs, we can now turn to the last requirement: existence of explicit estimation formulae.

The fact that permitted us to look for an inverse of a linear polynomial in

A_{II} among the set of linear polynomials in A_{II} is the degree of minimal polynomial of A_{II} : it is quadratic. In general, the inverse of any regular matrix P that is a polynomial $P(A)$, where A has a minimal polynomial of degree r, can be written as $Q(A)$, Q being of degree $r - 1$ at most.

Proof: The set of all such $Q(A)$ is a ring and in this ring the ideal generated by $P(A)$ must be the whole ring, otherwise it would be of lower dimension and $P(A)$ would be singular. Therefore, $P(A)$ has an inverse among the $Q(A)$.

In general for a I by I matrix r can be any number from 1 to I. As we have to find r constants in order to invert the matrix, the inversion can be done simply only for low r. The class $\mathbf{S}$ can be characterized as the class of matrices having the symmetries and optimality properties mentioned above, and a minimal polynomial of degree $r = 2$.

REFERENCES

[1] J. D. Bankier and R. E. Walpole, "Components of variance analysis for proportional frequencies", *Ann. Math. Stat.*, Vol. 28 (1957), pp. 742–53.

[2] C. Daniel, "Fractional replication of industrial experiments", *Transactions of the 1957 National Convention of the American Society for Quality Control*, pp. 229–233.

[3] J. D. Finney, "Main effects and interactions", *J. Amer. Stat. Assoc.*, Vol. 43 (1948), pp. 566–71.

[4] K. Kishen, "Symmetrical unequal block designs", *Sankhyā*, Vol. 5 (1941), pp. 329–348.

[5] K. R. Nair, "A note on the method of fitting constants for analysis of nonorthogonal data arranged in a double classification", *Sankhyā*, Vol. 5 (1941), pp. 317–28.

[6] H. Scheffé, *Analysis of Variance*, John Wiley and Sons, New York, 1959.

[7] G. W. Snedecor, *Statistical Methods*, 4th Edition, Iowa State College Press, Ames, Iowa, 1946.

[8] F. Yates, "The analysis of multiple classifications with unequal numbers in the different classes", *J. Amer. Stat. Assoc.*, Vol. 29 (1934), pp. 51–66.

A GENERALIZATION OF GROUP DIVISIBLE DESIGNS

By Damaraju Raghavarao

University of Bombay

1. Summary and Introduction. Roy [8] extended the idea of Group Divisible designs of Bose and Connor [1] to m-associate classes, calling such designs Hierarchical Group Divisible designs with m-associate classes. Subsequently, no literature is found in this direction. The purpose of this paper is to study these designs systematically. A compact definition of the design, under the name Group Divisible m-associate (GD m-associate) design is given in Section 2. In the same section the parameters of the design are obtained in a slightly different form than that of Roy. The uniqueness of the association scheme from the parameters is shown in Section 3. The designs are divided into $(m + 1)$ classes in Section 4. Some interesting combinatorial properties are obtained in Section 5. The necessary conditions for the existence of a class of these designs are obtained in Section 7. Finally, some numerical illustrations of these designs are given in the Appendix.

2. Definition and Parameters of a Group Divisible m-associate Design.

Definition 2.1. A Group Divisible m-associate design may be defined as follows:

(i) The experimental material is divided into b blocks of k units each, different treatments being applied to the units in the same block.

(ii) There are $v = N_1N_2 \cdots N_m$ treatments denoted by

$$v_{i_1 i_2 \cdots i_m}(i_1 = 1, 2, \cdots, N_1 ; i_2 = 1, 2, \cdots, N_2 ; \cdots ; i_m = 1, 2, \cdots, N_m).$$

Each treatment occurs once in each of the r blocks.

(iii) There can be established a relation of association between any two treatments satisfying the following requirements:

(a) Two treatments having only the first $(m - j)$ suffixes of $v_{i_1 i_2 \cdots i_m}$ the same are the jth associates $(j = 1, 2, \cdots, m)$.

(b) Each treatment has exactly n_j, jth associates.

(c) Given any two treatments which are ith associates, the number of treatments common to the jth associates of the first and the kth associates of the second is p^i_{jk} and is independent of the pair of treatments with which we start. Also, $p^i_{jk} = p^i_{kj}(i, j, k = 1, 2, \cdots, m)$.

(iv) Two treatments which are jth associates occur together in λ_j blocks.

The numbers $b, r, k, N_1, N_2, \cdots, N_m, \lambda_1, \lambda_2, \cdots, \lambda_m$ are known as the parameters of the GD m-associate design. We can easily see that

$$(2.1) \qquad n_i = N_m N_{m-1} \cdots N_{m-i+2}(N_{m-i+1} - 1), \qquad i = 1, 2, \cdots, m;$$

Received March 26, 1959; revised October 27, 1959.

and

$$(2.2) \qquad (p^i_{jk}) = \left[\begin{array}{c|c|c} 0_{(i-1)\times(i-1)} & x_{i-1} & 0_{(i-1)\times(m-i)} \\ \hline x'_{i-1} & \multicolumn{2}{c}{\multirow{2}{*}{$D_{(m-i+1)\times(m-i+1)}$}} \\ \cline{1-1} 0_{(m-i)\times(i-1)} & & \end{array}\right], \qquad i = 1, 2, \cdots, m,$$

where $0_{i\times i'}$ is a null matrix of the order $i \times i'$; x_{i-1} is the $(i-1)$th order column vector with elements $n_1, n_2, \cdots, n_{i-1}$; x'_{i-1} is the transpose of x_{i-1}; and

$$D_{(m-i+1)\times(m-i+1)}$$

is the diagonal matrix with elements $N_mN_{m-1}\cdots N_{m-i+2}(N_{m-i+1}-2)$, n_{i+1}, $n_{i+2}, \cdots, n_m$. The parameters satisfy the relations

$$N_1N_2\cdots N_mr = bk; \qquad \sum_{\alpha=1}^{m} n_\alpha = N_1N_2\cdots N_m - 1;$$

$$(2.3) \qquad \sum_{\alpha=1}^{m} n_\alpha\lambda_\alpha = r(k-1);$$

$$n_ip^i_{jk} = n_jp^j_{ik} = n_kp^k_{ij}; \qquad \sum_{k=1}^{m} p^i_{jk} = n_j - \delta_{ij}, \quad i, j, k = 1, 2, \cdots, m$$

where δ_{ij} is the Kronecker delta taking the value 1 or 0 according as $i = j$ or $i \neq j$. Since the parameters satisfy the above relations, it can be seen that a GD m-associate design is a special case of Partially Balanced Incomplete Block Designs defined by Bose and Nair [2].

3. Uniqueness of the Association Scheme. This section shows that the relations (2.1) and (2.2) imply the association scheme iii(a). In this section, we call a group of treatments which are first associates a first-associate group; a group of first-associate groups a second-associate group, etc. Let θ be any treatment. Let $\theta_1^{(i)}, \theta_2^{(i)}, \cdots, \theta_{n_i}^{(i)}$ be its ith associates $(i = 1, 2, \cdots, m)$. Consider the treatments θ and $\theta_1^{(1)}$. Since $n_1 = N_m - 1$ and $p^1_{11} = N_m - 2$, the first associates of $\theta_1^{(1)}$ except θ are the same as the first associates of θ except $\theta_1^{(1)}$. Also, as

$$p^1_{1i} = 0 (i = 2, 3, \cdots, m),$$

we can divide the treatments into first-associate groups such that treatments in different first-associate groups are 2nd, 3rd, $\cdots$, or mth associates. It can be seen that each first-associate group contains N_m treatments. Thus the v treatments are divided into $N_1N_2\cdots N_{m-1}$ first-associate groups of N_m treatments each.

Now, consider the treatments θ and $\theta_1^{(2)}$. Since

$$n_i = p^2_{ii} = N_mN_{m-1}\cdots N_{m-i+2}(N_{m-i+1}-1),$$

it is obvious that the ith associates of θ and $\theta_1^{(2)}$ are the same $(i = 3, 4, \cdots, m)$. Also, as $p^2_{11} = 0$ and $p^2_{22} = N_m(N_{m-1}-2)$, the $N_1N_2\cdots N_{m-1}$ first-associate groups of the above paragraph can be subdivided into $N_1N_2\cdots N_{m-2}$ second-

associate groups of N_{m-1} first-associate groups of N_m treatments each such that (i) treatments in different second-associate groups are 3rd, 4th, $\cdots$, or mth associates, and (ii) treatments in different first-associate groups of a second-associate group are the second associates.

Again, consider the treatments θ and $\theta_1^{(3)}$. Since

$$n_i = p_{ii}^3 = N_m N_{m-1} \cdots N_{m-i+2}(N_{m-i+1} - 1),$$

it can be seen that the ith associates of θ and $\theta_1^{(3)}$ are the same ($i = 4, 5, \cdots, m$). Also, as $p_{11}^3 = 0 = p_{12}^3 = p_{22}^3$ and $p_{23}^3 = N_m N_{m-1}(N_{m-2} - 2)$, the $N_1 N_2 \cdots N_{m-2}$ second-associate groups can be further grouped into $N_1 N_2 \cdots N_{m-3}$ third-associate groups each containing N_{m-2} second-associate groups. These second-associate groups contain N_{m-1} first-associate groups each containing N_m treatments. Treatments in different third-associate groups are 4th, 5th, $\cdots$, or mth associates. Treatments in different second-associate groups of a third-associate group are the third associates and treatments in different first-associate groups of a second-associate group are the second-associates.

By similar reasoning, we finally obtain N_1, $(m-1)$-associate groups of N_2, $(m-2)$-associate groups, $\cdots$, of N_{m-1} first-associate groups of N_m treatments. The above grouping will be such that (i) treatments in different $(m-1)$-associate groups are the mth associates, and (ii) treatments in different i-associate groups of an $(i+1)$-associate group are the $(i+1)$th associates

$$(i = 1, 2, \cdots, m-2).$$

We can easily see that the above grouping of the treatments is the same as the association scheme iii(a). Hence the parameters (2.1) and (2.2) define the association scheme iii(a) uniquely and we have the following:

THEOREM 3.1. *The relations* (2.1) *and* (2.2) *for a Group Divisible m-associate design uniquely define the association scheme* iii(a).

4. Characterization of Group Divisible m-associate Designs. Let $n_{ij} = 1$, if the ith treatment occurs in the jth block; and $n_{ij} = 0$, otherwise. Then the $v \times b$ matrix $N = (n_{ij})$ is known as the incidence matrix of the GD m-associate design. From the definition of GD m-associate design, we can see that

$$\sum_{j=1}^{b} n_{ij}^2 = r, \qquad i = 1, 2, \cdots, v; \quad \text{and} \quad \sum_{j=1}^{b} n_{ij} n_{i'j} = \lambda_1, \lambda_2, \cdots, \text{or } \lambda_m$$

according as i and i' are 1st, 2nd, $\cdots$, or mth associates, $i \neq i'$; $i, i' = 1, 2, \cdots, v$. Now, by suitably marking the treatments, we have

$$NN' = \begin{bmatrix} B_m & A_m & \cdots & A_m \\ A_m & B_m & \cdots & A_m \\ \vdots & \vdots & & \vdots \\ A_m & A_m & \cdots & B_m \end{bmatrix}, \tag{4.1}$$

where, at any stage,

$$B_i = \begin{bmatrix} B_{i-1} & A_{i-1} & \cdots & A_{i-1} \\ A_{i-1} & B_{i-1} & \cdots & A_{i-1} \\ \vdots & \vdots & & \vdots \\ A_{i-1} & A_{i-1} & \cdots & B_{i-1} \end{bmatrix}, \qquad i = 2, 3, \cdots, m;$$

(4.2)

$$A_i = \lambda_i E_{N_{m-i+2}N_{m-i+3}\cdots N_m}, \qquad i = 2, 3, \cdots, m;$$

$$B_1 = r,\ A_1 = \lambda_1,$$

where $E_{N_{m-i+2}N_{m-i+3}\cdots N_m}$, is an $N_{m-i+2}N_{m-i+3}\cdots N_m$th order square matrix with positive unit elements everywhere. The orders of NN' and B_i are $N_1N_2\cdots N_m$ and $N_{m-i+2}N_{m-i+3}\cdots N_m$ respectively $(i = 2, 3, \cdots, m)$. The matrices A_1 and B_1 are of unit order. Det (NN') can be evaluated in the usual manner and we get

$$|NN'| = rkP_1^{N_1-1}P_2^{N_1(N_2-1)}\cdots P_m^{N_1N_2\cdots N_{m-1}(N_m-1)}, \tag{4.3}$$

where

$$P_i = (r - \lambda_{m-i+1}) + (\lambda_1 - \lambda_{m-i+1})n_1 + \cdots + (\lambda_{m-i} - \lambda_{m-i+1})n_{m-i}, \qquad i = 1, 2, \cdots, m. \tag{4.4}$$

By replacing r by $(r - z)$ in det (NN') we can easily see that rk and P_i's $(i = 1, 2, \cdots, m)$ are the distinct characteristic roots of NN'. We know from the result of Connor and Clatworthy [4] that the characteristic roots of NN' cannot be negative for an existing design. Thus we have the following theorem:

THEOREM 4.1. *A necessary condition for the existence of a Group Divisible m-associate design is that* $P_i \geqq 0$ $(i = 1, 2, \cdots, m)$.

The designs with the following parameters violate the above necessary condition and hence are impossible. The reason of impossibility is shown in brackets against the parameters.

1. $v = 90 = b, r = 9 = k, N_1 = 3, N_2 = 15, N_3 = 2,$
$\lambda_1 = 12, \lambda_2 = 0, \lambda_3 = 1$ $(P_1, P_3 < 0)$.
2. $v = 12, b = 15, r = 5, k = 4, N_1 = 2 = N_2, N_3 = 3,$
$\lambda_1 = 0, \lambda_2 = 3, \lambda_3 = 1$ $(P_3 < 0)$.
3. $v = 8, b = 4, r = 3, k = 6, N_1 = 2 = N_2 = N_3,$
$\lambda_1 = 3, \lambda_2 = 0, \lambda_3 = 3$ $(P_1 < 0)$.
4. $v = 16 = b, r = 5 = k, N_1 = 2 = N_2, N_3 = 4,$
$\lambda_1 = 0, \lambda_2 = 1, \lambda_3 = 2$ $(P_1 < 0)$.
5. $v = 16, b = 24, r = 6, k = 4, N_1 = 2 = N_2 = N_3 = N_4,$
$\lambda_1 = 0, \lambda_2 = 1, \lambda_3 = 0, \lambda_4 = 2$ $(P_1 < 0)$.
6. $v = 32, b = 64, r = 10, k = 5, N_1 = 2 = N_2 = N_3 = N_4 = N_5,$
$\lambda_1 = 4, \lambda_2 = 0, \lambda_3 = 1, \lambda_4 = 0, \lambda_5 = 2$
$(P_1 < 0)$.

We can classify the existing designs mainly into $(m + 1)$ classes as follows:
(1) Singular GD m-associate designs characterized by $P_m = 0$;
(2) P_m − regular GD m-associate designs characterized by $P_m > 0, P_{m-1} = 0$;
.
.
.
(i) $P_m, P_{m-1}, \cdots, P_{m-i+2}$ − regular GD m-associate designs characterized by $P_m > 0, P_{m-1} > 0, \cdots, P_{m-i+2} > 0, P_{m-i+1} = 0$;
.
.
.
(m) $P_m, P_{m-1}, \cdots, P_2$ − regular GD m-associate designs characterized by $P_m > 0, P_{m-1} > 0, \cdots, P_2 > 0, P_1 = 0$; and
$(m + 1)$ Regular GD m-associate designs characterized by $P_i > 0$ $(i = 1, 2, \cdots m)$.
Excepting the last two classes, the other classes can be further divided; but, since this will be cumbersome, we do not do so.

5. Some Combinatorial Properties of Group Divisible m-associate Designs. If $P_i = 0 = P_{i+1} (i = 1, 2, \cdots, m - 1)$, we have $\lambda_{m-i+1} = \lambda_{m-i+2}$. Thus if $P_1 = 0 = P_2 = \cdots = P_m$, then $r = \lambda_1 = \cdots = \lambda_m$ and the GD m-associate design reduces to an ordinary randomised block design. Hence, we have

THEOREM 5.1. *If, in a Group Divisible m-associate design, $P_1 = 0 = P_2 = \cdots = P_m$, then the design reduces to a randomized block design.*

Let j consecutive λ's $(j = 2, 3, \cdots, m - 1)$ of the GD m-associate design be equal. In this case we can see from the association scheme that the design reduces to a GD $(m - j + 1)$-associate design. The above result can be written in the form of the following theorem.

THEOREM 5.2. *If, in a Group Divisible m-associate design j consecutive λ's $(j = 2, 3, \cdots, m - 1)$ are equal, then the design reduces to a Group Divisible $(m - j + 1)$-associate design.*

We now prove another important theorem.

THEOREM 5.3. *In a $P_m, P_{m-1}, \cdots, P_2$ − regular Group Divisible m-associate design k, is divisible by N_1. Further, every block contains k/N_1 treatments of the form $v_{ii_2\cdots i_m}$ $(i_2 = 1, 2, \cdots, N_2; i_3 = 1, 2, \cdots, N_3; \cdots; i_m = 1, 2, \cdots, N_m)$ for any i $(i = 1, 2, \cdots, N_1)$.*

PROOF. For any i $(i = 1, 2, \cdots, N_1)$, let e_j^i treatments of the form $v_{ii_2\cdots i_m}$ $(i_2 = 1, 2, \cdots, N_2; i_3 = 1, 2, \cdots, N_3; \cdots; i_m = 1, 2, \cdots, N_m)$ occur in the jth block $(j = 1, 2, \cdots, b)$. Then, we have

$$(5.1)\qquad \begin{aligned} \sum_{j=1}^{b} e_j^i &= N_2N_3 \cdots N_m r, \\ \sum_{j=1}^{b} e_j^i(e_j^i - 1) &= N_2N_3 \cdots N_m(n_1\lambda_1 + n_2\lambda_2 + \cdots + n_{m-1}\lambda_{m-1}), \end{aligned}$$

since each of the treatments occur in r blocks and every pair of treatments of the form $v_{ii_2\cdots i_m}$ $(i_2 = 1, 2, \cdots, N_2; i_3 = 1, 2, \cdots, N_3; \cdots; i_m = 1, 2, \cdots, N_m)$

occurs in $\lambda_1, \lambda_2, \cdots$, or λ_{m-1} blocks. Using the property of $P_m, P_{m-1}, \cdots, P_2$ – regular GD m-associate design and (5.1), we get

$$\sum_{j=1}^{b} (e_j^i)^2 = N_2^2 N_3^2 \cdots N_m^2 \lambda_m . \tag{5.2}$$

Let $\bar{e}^i = b^{-1} \sum_{j=1}^{b} e_j^i = k/N_1$. Then,

$$\sum_{j=1}^{b} (e_j^i - \bar{e}^i)^2 = N_2^2 N_3^2 \cdots N_m^2 \lambda_m - bk^2/N_1^2 = 0. \tag{5.3}$$

Therefore, $e_1^i = e_2^i = \cdots = e_b^i = \bar{e}^i = k/N_1$. Since $e_j^i (i = 1, 2, \cdots, N_1 ; j = 1, 2, \cdots, b)$ must be integral, k is divisible by N_1. Further $e_j^i = k/N_1$ $(i = 1, 2, \cdots, N_1 ; j = 1, 2, \cdots, b)$. This completes the proof of the theorem.

The following P_3, P_2-regular GD 3-associate designs have a non-integral value for k/N_1 and hence are non-existing:

1. $v = 12, b = 16, r = 4, k = 3, N_1 = 2, N_2 = 3, N_3 = 2,$
   $\lambda_1 = 2, \lambda_2 = 0, \lambda_3 = 1.$
2. $v = 12, b = 16, r = 4, k = 3, N_1 = 2 = N_2, N_3 = 3,$
   $\lambda_1 = 1, \lambda_2 = 0, \lambda_3 = 1.$
3. $v = 12, b = 9, r = 3, k = 4, N_1 = 3, N_2 = N_3 = 2,$
   $\lambda_1 = 1, \lambda_2 = 0, \lambda_3 = 1.$
4. $v = 20, b = 32, r = 8, k = 5, N_1 = 2, N_2 = 5, N_3 = 2,$
   $\lambda_1 = 4, \lambda_2 = 1, \lambda_3 = 2.$

A GD m-associate design is said to be symmetrical if $b = v$ and in consequence $r = k$. Shrikhande [9] and Chowla and Ryser [5] have obtained conditions necessary for the existence of symmetrical balanced incomplete block designs. Bose and Connor have obtained necessary conditions for the existence of symmetrical regular GD designs. We shall extend their results to symmetrical regular GD m-associate designs. With this in view, we give a brief resume of the important properties of the Legendre symbol, the Hilbert norm residue symbol and the Hasse-Minkowski invariant.

6. Some known results about the Legendre symbol, the Hilbert norm residue symbol and the Hasse-Minkowski invariant. The Legendre symbol is defined as

$$(a/p) = \begin{cases} +1, \text{ if } a \text{ is quadratic residue of } p; \\ -1, \text{ if } a \text{ is a non quadratic residue of } p. \end{cases} \tag{6.1}$$

A slight generalization of the Legendre symbol, is the Hilbert norm residue symbol $(a, b)_p$. If a and b are any non zero rational numbers, we define $(a, b)_p$ to have the value $+1$ or -1 according as the congruence

$$ax^2 + by^2 \equiv 1 \pmod{p^r}, \tag{6.2}$$

has or has not for every value of r, rational solutions x_r and y_r. Here p is any prime including the conventional prime $p_\infty = \infty$.

Many properties of $(a, b)_p$ are given by Bruck and Ryser [3], Jones [6] and

Pall [7]. For further use, we reproduce the properties of $(a, b)_p$ taken from the above references, in the form of the following theorems.

THEOREM 6.1. *If m and m' are integers not divisible by the odd prime p, then*

$$(m, m')_p = +1, \tag{6.3}$$

$$(m, p)_p = (m/p). \tag{6.4}$$

Moreover, if $m \equiv m' \not\equiv 0 \pmod p$, *then*

$$(m, p)_p = (m', p)_p. \tag{6.5}$$

THEOREM 6.2. *For arbitrary non-zero integers m, m', n, n', and for every prime p,*

$$(-m, m)_p = +1, \tag{6.6}$$

$$(m, n)_p = (n, m)_p, \tag{6.7}$$

$$(mm', n)_p = (m, n)_p(m', n)_p, \tag{6.8}$$

$$(m, nn')_p = (m, n)_p(m, n')_p, \tag{6.9}$$

$$(mm', m - m')_p = (m, -m')_p, \tag{6.10}$$

$$\prod_{j=1}^{m} (j, j + 1)_p = ((m + 1)!, -1)_p, \tag{6.11}$$

and

$$(as^2, b)_p = (a, b)_p. \tag{6.12}$$

Now, let $A = (a_{ij})$ be any $n \times n$ symmetric matrix with rational elements. The matrix B is said to be rationally congruent to A, written $A \sim B$, provided there exists a non-singular matrix C with rational elements, such that $A = CBC'$, where C' is the transpose of C. If D_i $(i = 1, 2, \cdots, n)$ denotes the leading principal minor determinant of order i in the matrix A, then if none of the D_i vanishes, the quantity

$$C_p = C_p(A) = (-1, -D_n)_p \prod_{i=1}^{n-1} (D_i, -D_{i+1})_p, \tag{6.13}$$

is invariant for all matrices rationally congruent to A. $C_p(A)$ defined above is known as the Hasse-Minkowski invariant.

The following lemmas regarding C_p will be useful.

LEMMA 6.1. *If d is a rational number and $\Delta_m = dI_m$, where I_m is the identity matrix of order m, then*

$$C_p(\Delta_m) = (-1, -1)_p(d, -1)_p^{m(m+1)/2}. \tag{6.14}$$

LEMMA 6.2. *If A and B are symmetric matrices with rational elements and $U = A \dotplus B$, is the direct sum of A and B, then*

$$C_p(U) = (-1, -1)_p C_p(A)C_p(B)(|A|, |B|)_p. \tag{6.15}$$

7. Necessary conditions for the existence of Symmetrical Regular Group Divisible m-associate Designs. Since the design is a symmetric one, $\det(NN')$ is a perfect square (cf. Connor and Clatworthy, and Shrikhande). Thus

$$P_1^{N_1-1} P_2^{N_1(N_2-1)} \cdots P_m^{N_1 N_2 \cdots N_{m-1}(N_m-1)}$$

is a perfect square. This result can be written in the form of the following theorem.

THEOREM 7.1. *A necessary condition for the existence of a regular symmetrical Group Divisible m-associate design is that* $P_1^{N_1-1} P_2^{N_1(N_2-1)} \cdots P_m^{N_1 N_2 \cdots N_{m-1}(N_m-1)}$ *is a perfect square.*

The designs with the following parameters do not satisfy the above theorem and hence are impossible.

1. $v = 24 = b, r = 6 = k, N_1 = 4, N_2 = 2, N_3 = 3,$
$\lambda_1 = 3, \lambda_2 = 2, \lambda_3 = 1.$
2. $v = 32 = b, r = 7 = k, N_1 = 4, N_2 = 2, N_3 = 4,$
$\lambda_1 = 2, \lambda_2 = 3, \lambda_3 = 1.$
3. $v = 30 = b, r = 7 = k, N_1 = 5, N_2 = 3, N_3 = 2,$
$\lambda_1 = 2, \lambda_2 = 4, \lambda_3 = 1.$
4. $v = 30 = b, r = 12 = k, N_1 = 3, N_2 = 5, N_3 = 2,$
$\lambda_1 = 4, \lambda_2 = 6, \lambda_3 = 4.$
5. $v = 54 = b, r = 11 = k, N_1 = 3 = N_2 = N_3, N_4 = 2,$
$\lambda_1 = 6, \lambda_2 = 5, \lambda_3 = 4, \lambda_4 = 1.$

Let

$$Q_i = B_i - A_i, \qquad i = 2, 3, \cdots, m; \tag{7.1}$$

where B_i's and A_i's are as defined in Section 4. $\operatorname{Det}(Q_i) i = 2, 3, \cdots, m$ can be found easily, and we have

$$\begin{aligned} |Q_i| = &\{(r - \lambda_i) + (\lambda_1 - \lambda_i)n_1 + \cdots + (\lambda_{i-1} - \lambda_i)n_{i-1}\} \\ &\{(r - \lambda_{i-1}) + (\lambda_1 - \lambda_{i-1})n_1 + \cdots + (\lambda_{i-2} - \lambda_{i-1})n_{i-2}\}^{N_{m-i+2}-1} \\ &\cdot \\ &\cdot \\ &\cdot \\ &\{r - \lambda_1\}^{N_{m-i+2}N_{m-i+3}\cdots N_{m-1}(N_m-1)}. \end{aligned} \tag{7.2}$$

Now, let us calculate the Hasse-Minkowski invariant of (NN') for odd primes using the method of Bose and Connor. Taking the direct sum with $-\lambda_m$, NN' becomes

$$(NN')_1 = \begin{bmatrix} NN' & \\ & -\lambda_m \end{bmatrix}. \tag{7.3}$$

Therefore, from Lemma 6.2,

$$C_p(NN')_1 = C_p(NN')(\lambda_m, -1)_p. \tag{7.4}$$

But

$$(NN')_1 \sim \begin{bmatrix} Q_m & & & & L \\ & Q_m & & & L \\ & & \ddots & & \vdots \\ & & & Q_m & L \\ L & L & \cdots & L & -\lambda_m \end{bmatrix} \tag{7.5}$$

where L is an $N_2N_3 \cdots N_m$th order column vector with $-\lambda_m$ everywhere. Hence

$$C_p(NN')_1 = \{C_p(Q_m)\}^{N_1}(|Q_m|, -1)_p^{N_1(N_1+1)/2}(\lambda_m, -|Q_m|^{N_1})_p. \tag{7.6}$$

Equating (7.4) and (7.6), we get

$$C_p(NN') = \{C_p(Q_m)\}^{N_1}(|Q_m|, -1)_p^{N_1(N_1+1)/2}(\lambda_m, |Q_m|)_p^{N_1}. \tag{7.7}$$

$C_p(Q_i)i = 2, 3, \cdots, m$ can be calculated in a similar way as above and we get

$$\begin{aligned} C_p(Q_2) = {} & (r - \lambda_1, -1)_p^{N_m(N_m-1)/2}(\lambda_1 - \lambda_2, r - \lambda_1)_p^{N_m} \\ & (|Q_2|, r - \lambda_1)_p^{N_m}(|Q_2|, \lambda_1 - \lambda_2)_p, \end{aligned} \tag{7.8}$$

$$\begin{aligned} C_p(Q_i) = {} & \{C_p(Q_{i-1})\}^{N_{m-i+2}}(|Q_{i-1}|, -1)_p^{N_{m-i+2}(N_{m-i+2}+1)/2} \\ & (\lambda_{i-1} - \lambda_i, |Q_i|)_p(\lambda_{i-1} - \lambda_i, |Q_{i-1}|)^{N_{m-i+2}} \\ & (|Q_i|, |Q_{i-1}|)^{N_{m-i+2}}, \qquad i = 3, 4, \cdots, m. \end{aligned} \tag{7.9}$$

Equation (7.9) is a recurrence relation. This equation with the help of (7.2) and (7.8) finally gives $C_p(Q_m)$. Substituting this value of $C_p(Q_m)$ in (7.7), $C_p(NN')$ can be calculated. Now, since $I_v = N^{-1}(NN')(N')^{-1}$, $I_v \sim NN'$. Therefore,

$$C_p(NN') = C_p(I_v) = (-1, -1)_p = +1. \tag{7.10}$$

Thus we have the following theorem

THEOREM 7.2. *A necessary condition for the existence of a symmetrical regular Group Divisible m-associate design is that* $C_p(NN') = +1$, *for odd primes p where* $C_p(NN')$ *is calculated from* (7.2), (7.8), (7.9) *and* (7.7).

When there are only three associate classes the above calculations can be simplified and the corollary follows:

COROLLARY 7.2.1. *A necessary condition for the existence of a regular symmetrical Group Divisible 3-associate design is that*

$$\begin{aligned} & (P_3, -1)_p^{\{N_1N_2N_3(N_1+N_2+N_3+3)-N_1N_2(N_1+N_2)\}/2}(\lambda_3, P_1)_p \\ & \cdot (P_1, -1)_p^{N_1(N_1+1)/2}(P_2, -1)^{N_1\{N_2(N_2+1)+(N_2-1)(N_1+1)\}/2} \\ & \cdot (\lambda_2 - \lambda_3, P_1P_2)_p^{N_1}(\lambda_1 - \lambda_2, P_2P_3)_p^{N_1N_2}(P_2, P_3)_p^{N_1N_2} \\ & \cdot (P_1, P_2)_p^{N_1N_2}(P_1, P_3)_p^{N_1N_2(N_2-1)} = +1, \qquad \text{for all odd primes } p. \end{aligned} \tag{7.11}$$

ILLUSTRATION 7.2.1. Consider the GD 3-associate design with the parameters

$$v = 27 = b, r = 7 = k, N_1 = 3 = N_2 = N_3, \lambda_1 = 6, \lambda_2 = 2, \lambda_3 = 1.$$

The left hand side of (7.11) is

$$(22, 13)_p = (13, 2)_p(13, 11)_p = -1, \qquad \text{when } p = 11.$$

Thus the corollary 7.2.1 is not satisfied and the design is impossible.

ILLUSTRATION 7.2.2. For the GD 3-associate design with the parameters

$$v = 48 = b, r = 10 = k, N_1 = 6, N_2 = 4, N_3 = 2,$$
$$\lambda_1 = 4, \lambda_2 = 1, \lambda_3 = 2.$$

the left hand side of (7.11) is

$$(12, -1)_p = (3, -1)_p = -1, \text{ for } p = 3.$$

The Corollary 7.2.1 is not satisfied and the design is impossible.

By applying the Corollary 7.2.1, it can be easily verified that the following designs are non-existing:

1. $v = 24 = b, r = 9 = k, N_1 = 2 = N_2, N_3 = 6, \lambda_1 = 6, \lambda_2 = 1, \lambda_3 = 3.$
2. $v = 24 = b, r = 10 = k, N_1 = 2 = N_2, N_3 = 6, \lambda_1 = 6, \lambda_2 = 2, \lambda_3 = 4.$
3. $v = 24 = b, r = 10 = k, N_1 = 6, N_2 = 2 = N_3, \lambda_1 = 6, \lambda_2 = 2, \lambda_3 = 4.$
4. $v = 40 = b, \ r = 13 = k, \ N_1 = 10, \ N_2 = 2 = N_3, \ \lambda_1 = 10, \ \lambda_2 = 1, \lambda_3 = 4.$

8. Acknowledgment. My sincere thanks are due to Professor M. C. Chakrabarti for his kind guidance.

REFERENCES

[1] R. C. BOSE AND W. S. CONNOR, "Combinatorial properties of group divisible incomplete block designs," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 367–383.

[2] R. C. BOSE AND K. R. NAIR, "Partially balanced incomplete block designs," *Sankhyā*, Vol. 4 (1939), pp. 337–372.

[3] R. H. BRUCK AND H. J. RYSER, "The non existence of certain finite projective planes," *Can. J. Math.*, Vol. 1 (1949), pp. 88–93.

[4] W. S. CONNOR AND W. H. CLATWORTHY, "Some theorems for partially balanced designs," *Ann. Math., Stat.*, Vol. 25 (1954), pp. 100–112.

[5] S. CHOWLA AND H. J. RYSER, "Combinatorial problems," *Can. J. of Math.*, Vol. 2 (1950), pp. 93–99.

[6] B. W. JONES, *The Arithmetical Theory of Quadratic forms*, John Wiley & Sons, New York, 1950.

[7] G. PALL, "The arithmetical invariant of quadratic forms," *Bull. Am. Math. Soc.*, Vol. 51 (1945), pp. 185–197.

[8] PURNENDU MOHON ROY, "Hierarchical group divisible incomplete block designs with m-associate classes," *Science and Culture*, Vol. 19 (1953–54), pp. 210–211.

[9] S. S. SHRIKHANDE, "The impossibility of certain symmetrical balanced incomplete block designs," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 106–111.

APPENDIX

Here we give some numerical constructions in the useful range $r,\ k \leqq 10$. For convenience we denote the treatment v_{ijk} by (ijk) in the following examples.

1. $v = 8 = b$, $r = 3 = k$, $N_1 = 2 = N_2 = N_3$, $\lambda_1 = 2$, $\lambda_2 = 0$, $\lambda_3 = 1$. Taking the treatments as

(111)	(112)	(211)	(212)
(121)	(122)	(221)	(222)

the plan of the design is

	[(111)	(112)	(211)]
	[(112)	(111)	(212)]
	[(121)	(122)	(221)]
	[(122)	(121)	(222)]
	[(211)	(212)	(121)]
	[(212)	(211)	(122)]
	[(221)	(222)	(111)]
	[(222)	(221)	(112)]
Reps.	I	II	III

2. $v = 8 = b$, $r = 4 = k$, $N_1 = 2 = N_2 = N_3$, $\lambda_1 = 2$, $\lambda_2 = 1$, $\lambda_3 = 2$. Taking the treatments as in the above example, the plan of the design is

	[(111)	(112)	(211)	(221)]
	[(112)	(111)	(212)	(222)]
	[(121)	(122)	(221)	(212)]
	[(122)	(121)	(222)	(211)]
	[(222)	(221)	(111)	(121)]
	[(221)	(222)	(112)	(122)]
	[(211)	(212)	(122)	(111)]
	[(212)	(211)	(121)	(112)]
Reps.	I	II	III	IV

3. $v = 8$, $b = 24$, $r = 9$, $k = 3$, $N_1 = 2 = N_2 = N_3$, $\lambda_1 = 4$, $\lambda_2 = 1$, $\lambda_3 = 3$. Taking the treatments as in Example 1, the plan of the design is

			Reps.
[(111)	(112)	(211)]	
[(112)	(111)	(212)]	
[(121)	(122)	(221)]	
[(122)	(121)	(222)]	I, II, III
[(211)	(212)	(121)]	
[(212)	(211)	(122)]	
[(221)	(222)	(111)]	
[(222)	(221)	(112)]	

[(111)	(112)	(221)]	
[(112)	(111)	(222)]	
[(121)	(122)	(211)]	
[(122)	(121)	(212)]	IV, V, VI
[(211)	(212)	(111)]	
[(212)	(211)	(112)]	
[(221)	(222)	(121)]	
[(222)	(221)	(122)]	
[(111)	(121)	(211)]	
[(122)	(111)	(221)]	
[(121)	(112)	(222)]	
[(112)	(122)	(212)]	VII, VIII, IX
[(211)	(221)	(112)]	
[(222)	(211)	(122)]	
[(221)	(212)	(121)]	
[(212)	(222)	(111)]	
Reps. I, II, III	IV, V, VI	VII, VIII, IX	

4. $v = 8 = b,\ r = 5 = k,\ N_1 = 2 = N_2 = N_3,\ \lambda_1 = 4,\ \lambda_2 = 2,\ \lambda_3 = 3.$ Taking the treatments as in Example 1, the plan of the design is

	[(111)	(112)	(211)	(221)	(222)]
	[(112)	(111)	(212)	(222)	(221)]
	[(121)	(122)	(222)	(211)	(212)]
	[(122)	(121)	(221)	(212)	(211)]
	[(211)	(212)	(121)	(111)	(112)]
	[(212)	(211)	(122)	(112)	(111)]
	[(221)	(222)	(111)	(121)	(122)]
	[(222)	(221)	(112)	(122)	(121)]
Reps.	I	II	III	IV	V

5. $v = 8 = b,\ r = 6 = k,\ N_1 = 2 = N_2 = N_3,\ \lambda_1 = 4,\ \lambda_2 = 5,\ \lambda_3 = 4.$ Taking the treatments as in Example 1, the plan of the design is

	[(111)	(112)	(121)	(122)	(211)	(221)]
	[(112)	(121)	(122)	(111)	(222)	(211)]
	[(121)	(122)	(111)	(112)	(221)	(212)]
	[(122)	(111)	(112)	(121)	(212)	(222)]
	[(211)	(212)	(221)	(222)	(111)	(121)]
	[(212)	(221)	(222)	(211)	(122)	(111)]
	[(221)	(222)	(211)	(212)	(112)	(122)]
	[(222)	(211)	(212)	(221)	(121)	(112)]
Reps.	I	II	III	IV	V	VI

6. $v = 12 = b,\ r = 4 = k,\ N_1 = 2 = N_2,\ N_3 = 3,\ \lambda_1 = 3,\ \lambda_2 = 0,\ \lambda_3 = 1.$

Taking the treatments as

(111)	(112)	(113)	(211)	(212)	(213)
(121)	(122)	(123)	(221)	(222)	(223)

the plan of the design is

	[(111)	(112)	(113)	(211)]
	[(112)	(113)	(111)	(212)]
	[(113)	(111)	(112)	(213)]
	[(121)	(122)	(123)	(221)]
	[(122)	(123)	(121)	(222)]
	[(123)	(121)	(122)	(223)]
	[(211)	(212)	(213)	(121)]
	[(212)	(213)	(211)	(122)]
	[(213)	(211)	(212)	(123)]
	[(221)	(222)	(223)	(111)]
	[(222)	(223)	(221)	(112)]
	[(223)	(221)	(222)	(113)]
Reps.	I	II	III	IV

7. $v = 16 = b$, $r = 4 = k$, $N_1 = 2 = N_2 = N_3 = N_4$, $\lambda_1 = 0$, $\lambda_2 = 2$, $\lambda_3 = 0$, $\lambda_4 = 1$. Taking the treatments as

(1111)	(1112)	(1211)	(1212)
(1121)	(1122)	(1221)	(1222)
(2111)	(2112)	(2211)	(2212)
(2121)	(2122)	(2221)	(2222)

the plan of the design is

	[(1111)	(1121)	(2111)	(2121)]
	[(1121)	(1111)	(2112)	(2122)]
	[(2211)	(2221)	(1111)	(1122)]
	[(2212)	(2222)	(1122)	(1111)]
	[(2221)	(2211)	(1112)	(1121)]
	[(2222)	(2212)	(1121)	(1112)]
	[(1112)	(1122)	(2121)	(2111)]
	[(1122)	(1112)	(2122)	(2112)]
	[(2111)	(2122)	(1211)	(1222)]
	[(2112)	(2121)	(1222)	(1211)]
	[(1211)	(1221)	(2211)	(2222)]
	[(1221)	(1211)	(2212)	(2221)]
	[(1222)	(1212)	(2222)	(2211)]
	[(1212)	(1222)	(2221)	(2212)]
	[(2122)	(2111)	(1221)	(1212)]
	[(2121)	(2112)	(1212)	(1221)]
Reps.	I	II	III	IV

8. $v = 16$, $b = 32$, $r = 8$, $k = 4$, $N_1 = 2 = N_2 = N_3 = N_4$, $\lambda_1 = 4$, $\lambda_2 = 2$, $\lambda_3 = 0$, $\lambda_4 = 2$. Taking the treatments as in the above example, the plan of the design is

					Reps.
	[(1111)	(1112)	(2111)	(2121)]	
	[(1112)	(1111)	(2112)	(2122)]	
	[(2111)	(2122)	(1121)	(1122)]	
	[(2112)	(2121)	(1122)	(1121)]	I, II
	[(1211)	(1212)	(2211)	(2221)]	
	[(1212)	(1211)	(2212)	(2222)]	
	[(2211)	(2222)	(1221)	(1222)]	
	[(2212)	(2221)	(1222)	(1221)]	
	[(2221)	(2211)	(1111)	(1112)]	
	[(2222)	(2212)	(1112)	(1111)]	
	[(1121)	(1122)	(2222)	(2211)]	
	[(1122)	(1121)	(2221)	(2212)]	III, IV
	[(2121)	(2111)	(1211)	(1212)]	
	[(2122)	(2112)	(1212)	(1211)]	
	[(1221)	(1222)	(2122)	(2111)]	
	[(1222)	(1221)	(2121)	(2112)]	
	[(2111)	(2112)	(1111)	(1121)]	
	[(2112)	(2111)	(1112)	(1122)]	
	[(1111)	(1122)	(2121)	(2122)]	
	[(1112)	(1121)	(2122)	(2121)]	V, VI
	[(2211)	(2212)	(1211)	(1221)]	
	[(2212)	(2211)	(1212)	(1222)]	
	[(1211)	(1222)	(2221)	(2222)]	
	[(1212)	(1221)	(2222)	(2221)]	
	[(1121)	(1111)	(2211)	(2212)]	
	[(1122)	(1112)	(2212)	(2211)]	
	[(2221)	(2222)	(1122)	(1111)]	
	[(2222)	(2221)	(1121)	(1112)]	VII, VIII
	[(1221)	(1211)	(2111)	(2112)]	
	[(1222)	(1212)	(2112)	(2111)]	
	[(2121)	(2122)	(1222)	(1211)]	
	[(2122)	(2121)	(1221)	(1212)]	
Reps.	I, II	III, IV	V, VI	VII, VIII	

9. $v = 18$, $b = 42$, $r = 7$, $k = 3$, $N_1 = 3$, $N_2 = 3$, $N_3 = 2$, $\lambda_1 = 2$, $\lambda_2 = 0$,

$\lambda_3 = 1$. Taking the treatments as

(111)	(112)	(211)	(212)	(311)	(312)
(121)	(122)	(221)	(222)	(321)	(322)
(131)	(132)	(231)	(232)	(331)	(332)

the plan of the design is

[(111), (112), (211)]	[(111), (112), (212)]
[(121), (122), (221)]	[(121), (122), (222)]
[(131), (132), (231)]	[(131), (132), (232)]
[(211), (212), (311)]	[(211), (212), (312)]
[(221), (222), (321)]	[(221), (222), (322)]
[(231), (232), (331)]	[(231), (232), (332)]
[(311), (312), (111)]	[(311), (312), (112)]
[(321), (322), (121)]	[(321), (322), (122)]
[(331), (332), (131)]	[(331), (332), (132)]
[(111), (221), (331)]	[(111), (222), (332)]
[(111), (231), (321)]	[(111), (232), (322)]
[(112), (221), (332)]	[(112), (222), (331)]
[(112), (231), (322)]	[(112), (232), (321)]
[(121), (211), (331)]	[(121), (212), (332)]
[(121), (231), (311)]	[(121), (232), (312)]
[(122), (211), (332)]	[(122), (232), (311)]
[(122), (231), (312)]	[(122), (212), (331)]
[(131), (211), (321)]	[(131), (212), (322)]
[(131), (221), (311)]	[(131), (222), (312)]
[(132), (211), (322)]	[(132), (212), (321)]
[(132), (221), (312)]	[(131), (222), (311)]

10. $v = 24, b = 16, r = 4, k = 6, N_1 = 3, N_2 = 4, N_3 = 2, \lambda_1 = 4, \lambda_2 = 0,$ $\lambda_3 = 1$. Taking the treatments as

(111)	(112)	(211)	(212)	(311)	(312)
(121)	(122)	(221)	(222)	(321)	(322)
(131)	(132)	(231)	(232)	(331)	(332)
(141)	(142)	(241)	(242)	(341)	(342)

the plan of the design is

[(111), (112), (211), (212), (311), (312)]
[(111), (112), (221), (222), (321), (322)]
[(111), (112), (231), (232), (331), (332)]
[(111), (112), (241), (242), (341), (342)]
[(121), (122), (211), (212), (321), (322)]
[(121), (122), (221), (222), (331), (332)]
[(121), (122), (231), (232), (341), (342)]
[(121), (122), (241), (242), (311), (312)]

[(131), (132), (211), (212), (331), (332)]
[(131), (132), (221), (222), (341), (342)]
[(131), (132), (231), (232), (311), (312)]
[(131), (132), (241), (242), (321), (322)]
[(141), (142), (211), (212), (341), (342)]
[(141), (142), (221), (222), (311), (312)]
[(141), (142), (231), (232), (321), (322)]
[(141), (142), (241), (242), (331), (332)]

RELATIONS AMONG THE BLOCKS OF THE KRONECKER PRODUCT OF DESIGNS

By Manohar Narhar Vartak

University of Bombay

1. Summary and Introduction. In the case of some incomplete block designs, interesting relations among their blocks have been discovered. For example, Fisher [1] has shown that in the case of a symmetrical BIB (Balanced Incomplete Block) design with parameters $v = b$, $r = k$, λ, any two blocks have exactly λ treatments in common. Similarly, Bose [2] has shown that in the case of an affine resolvable BIB design with parameters

$$v = nk = n^2\{(n-1)t+1\}, \qquad b = nr = n\{n^2t+n+1\}, \qquad \lambda = nt+1,$$

the blocks can be divided into sets of n blocks, such that each set is a complete replication and any two blocks have $(k^2)/v = (nt - t + 1)$ or 0 treatments in common according as they belong to different groups or the same group. Also see Connor [3] and Bose and Connor [4] for similar results.

Confining our attention to PBIB (Partially Balanced Incomplete Block) designs with two or three associate classes, we wish to see how this type of information for blocks of BIB designs can be used to obtain similar information for the blocks of their Kronecker product.

In the next section are given a few general properties of the Kronecker product of designs. In Section 3 the main theorems of the paper are proved and their important particular cases are discussed. Some observations on the interconnection between these results and the theorems on inversion of designs (cf. Roy [5], Shrikhande [6]) are made in Section 4.

2. Some general properties of the Kronecker product of designs. We shall always denote the Kronecker product of matrices A and B by $A \times B$ (cf. Vartak [7]); and the ordinary product of A and B, whenever it exists, will be denoted by $A \cdot B$ or AB. The Kronecker product of designs was defined in [7] as the design whose incidence matrix is the Kronecker product of the incidence matrices of the given designs.

We shall consider throughout this section two designs N_1 and N_2 with v_1 and v_2 treatments and b_1 and b_2 blocks respectively. The design N_1' whose incidence matrix N_1' is the transpose of N_1, is said to be the design obtained from N_1 by inversion [5], or dualization [6]. Similarly for the design N_2'. Since the Kronecker product of matrices satisfies the law

$$(A \times B)' = A' \times B' \tag{2.1}$$

we get the following result for the inversion of the Kronecker product of designs.

Theorem 2.1. *The design obtained by the inversion of the Kronecker product of*

Received September 8, 1959; revised December 8, 1959.

two given designs is the same as the Kronecker product of the inversions of the given designs. Thus if N_1 and N_2 are both symmetric (or self-dual), their Kronecker product is also symmetric (or self-dual).

In many cases we are interested in the matrix NN' where N is the incidence matrix of a given design. Let $N = N_1 \times N_2$, where N_1 and N_2 are the given designs. Clearly the Kronecker product of matrices satisfies the relation

$$(AB) \times (CD) = (A \times C) \cdot (B \times D) \tag{2.2}$$

where A, B, C and D are matrices of orders $m \times k$, $k \times n$, $p \times j$ and $j \times q$ respectively. Both sides of (2.2) are then $mp \times nq$ matrices. Hence we get

THEOREM 2.2. *The matrix NN' for the Kronecker product $N = N_1 \times N_2$ of two given designs is the Kronecker product of the corresponding matrices for the given designs; similarly for the matrix $N'N$.*

Finally, we need the following two results from [7].

2A. The Kronecker product $N = N_1(\text{BIB}) \times N_2(\text{BIB})$ of two BIB designs $N_1(\text{BIB})$ and $N_2(\text{BIB})$ defined by the respective sets of parameters

$$v_1, b_1, r_1, k_1, \lambda_1 \tag{2.3}$$

and

$$v_2, b_2, r_2, k_2, \lambda_2 \tag{2.4}$$

is a PBIB design with at most three associate classes.

The three associate classes of the design N defined above are all distinct if $r_1\lambda_2 \neq r_2\lambda_1$.

In any case, the parameters of the design N can be expressed in terms of those of the BIB designs given by (2.3) and (2.4) by the following equations

$$\begin{gathered}
v' = v_1v_2, \quad b' = b_1b_2, \quad r' = r_1r_2, \quad k' = k_1k_2, \\
n_1' = v_2 - 1, \quad n_2' = v_1 - 1, \quad n_3' = n_1n_2, \\
\lambda_1' = r_1\lambda_2, \quad \lambda_2' = r_2\lambda_1, \quad \lambda_3' = \lambda_1\lambda_2, \\
(p_{yz}'^1) = \begin{bmatrix} v_2 - 2 & 0 & 0 \\ 0 & 0 & v_1 - 1 \\ 0 & v_1 - 1 & (v_1 - 1)(v_2 - 2) \end{bmatrix}, \\
(p_{yz}'^2) = \begin{bmatrix} 0 & 0 & v_2 - 1 \\ 0 & v_1 - 2 & 0 \\ v_2 - 1 & 0 & (v_1 - 2)(v_2 - 1) \end{bmatrix}, \\
(p_{yz}'^3) = \begin{bmatrix} 0 & 1 & v_2 - 2 \\ 1 & 0 & v_1 - 2 \\ v_2 - 2 & v_1 - 2 & (v_1 - 2)(v_2 - 2) \end{bmatrix}
\end{gathered} \tag{2.5}$$

where $y, z = 1, 2, 3$.

As a direct consequence of 2A and Theorems 2.1 and 2.2, we get the following corollary

Corollary 2.1.1. *If a symmetrical* PBIB *design (i.e., one with $v = b$ and hence $r = k$) with three associate classes and parameters* (2.5) *is the Kronecker product of two symmetrical* BIB *designs, then with respect to any block B in it, the other blocks fall into three groups (α), (β) and (γ) such that the group (α) contains n_1' blocks each having λ_1' treatments in common with B, the group (β) contains n_2' blocks each having λ_2' treatments in common with B, and the group (γ) contains n_3' blocks each having λ_3' treatments in common with B.*

Proof. Let the given symmetrical PBIB design N be the Kronecker product of the symmetrical BIB designs N_1(BIB) and N_2(BIB) with respective sets of parameters

$$v_1 = b_1, \qquad r_1 = k_1, \qquad \lambda_1$$

and

$$v_2 = b_2, \qquad r_2 = k_2, \qquad \lambda_2.$$

By the well known result in [1], it follows that

$$N_1'(\text{BIB})\cdot N_1(\text{BIB}) = (r_1 - \lambda_1)I_{v_1} + \lambda_1 E_{v_1 v_1}$$

where I_{v_1} is the identity matrix of order v_1 and $E_{v_1 v_1}$ is the matrix of order $v_1 \times v_1$ with all elements equal to 1. Similarly for the design N_2(BIB). Since

$$N = N_1(\text{BIB}) \times N_2(\text{BIB}),$$

it follows from Theorem 2.2 that

$$N'N = \{(r_1 - \lambda_1)I_{v_1} + \lambda_1 E_{v_1 v_1}\} \times \{(r_2 - \lambda_2)I_{v_2} + \lambda_2 E_{v_2 v_2}\},$$

which, in virtue of (2.5), simplifies to

$$N'N = \begin{bmatrix} A & B & \cdots & B \\ B & A & \cdots & B \\ \cdots & \cdots & \cdots & \cdots \\ B & B & \cdots & A \end{bmatrix}$$

where

$$A = (r' - \lambda_1')I_{v_2} + \lambda_1' E_{v_2 v_2}$$

and

$$B = (\lambda_2' - \lambda_3')I_{v_2} + \lambda_3' E_{v_2 v_2}.$$

The result of Corollary 2.1.1 follows from the fact that the element in the ith row and the jth column of $N'N$ equals the number of treatments common to the ith and the jth blocks.

2B. A set of necessary and sufficient conditions for the Kronecker product N of the two BIB designs given by (2.3) and (2.4) to have only two distinct associate classes is given by

$$(2.6) \qquad v_1 = v_2 = v \quad \text{say}, \qquad \text{and} \quad k_1 = k_2 = k \quad \text{say}.$$

If these conditions are fulfilled, then it follows that

$$b_2/b_1 = r_2/r_1 = \lambda_2/\lambda_1 = \mu, \quad \text{say} \tag{2.7}$$

where μ is a positive fraction; and in this case the parameters of N can be expressed in terms of those of the BIB designs by the equations

$$\begin{gathered} v' = v^2, \quad b' = \mu b_1^2, \quad r' = \mu r_1^2, \quad k' = k^2, \\ n_1' = 2(v-1), \quad n_2' = (v-1)^2, \\ \lambda_1' = \mu r_1 \lambda_1, \quad \lambda_2' = \mu \lambda_1^2, \\ (p_{yz}'^1) = \begin{bmatrix} v-2 & v-1 \\ v-1 & (v-1)(v-2) \end{bmatrix}, \quad (p_{yz}'^2) = \begin{bmatrix} 2 & 2(v-2) \\ 2(v-2) & (v-2)^2 \end{bmatrix}, \end{gathered} \tag{2.8}$$

where $y, z = 1, 2$.

Both the results 2A and 2B are particular cases of a general result, Theorem 4.2 of [7].

3. The Main Theorems. Let N be the incidence matrix of a given design with parameters v, b, r, k. Then the matrix

$$N'N = (n_{ij}'); \quad i, j = 1, 2, \cdots, b; \tag{3.1}$$

is such that its general element n_{ij}' gives the number of treatments common to the ith and the jth blocks of N. If N_1 and N_2 are two designs with parameters v_1, b_1, r_1, k_1 and v_2, b_2, r_2, k_2 respectively, then the matrix $N'N$ for the Kronecker product $N = N_1 \times N_2$ is given by

$$N'N = (N_1'N_1) \times (N_2'N_2). \tag{3.2}$$

From this we get the following theorem.

THEOREM 3.1. *If in the design N_1 there exists a pair of blocks having m_1 treatments in common and in the design N_2 a pair of blocks having m_2 treatments in common, then in their Kronecker product $N = N_1 \times N_2$ there exists a pair of blocks having m_1m_2 treatments in common.*

PROOF. It is clear that m_1 will be an element of $N_1'N_1$ and m_2 of $N_2'N_2$; so that by (3.2) $N'N$ will contain m_1m_2 as an element. This proves Theorem 3.1.

Now consider a block $B^{(1)}$ of the design N_1 and let $b_i^{(1)}$ of the totality of the blocks of N_1 have each i treatments in common with $B^{(1)}$; $i = 0, 1, 2, \cdots, k_1$. Clearly $\sum_{i=0}^{k_1} b_i^{(1)} = b_1$. Let $B^{(2)}$ and $b_j^{(2)}$ have similar meanings for the design N_2 so that $\sum_{j=0}^{k_2} b_j^{(2)} = b_2$. Remembering that the blocks of N_1 are of size k_1 and those of N_2 are of size k_2 we get the following theorem.

THEOREM 3.2. *If there exist blocks $B^{(1)}$ and $B^{(2)}$ in the designs N_1 and N_2 respectively having the above properties, then there exists a block B in the Kronecker product $N = N_1 \times N_2$ such that $b_u^{(0)}$ blocks of N have each u treatments in common with B, where $b_u^{(0)}$ is the coefficient of $\{u\}$ in the expression*

$$\left(\sum_{i=0}^{k_1} b_i^{(1)}\{i\}\right)\left(\sum_{j=0}^{k_2} b_j^{(2)}\{j\}\right) \tag{3.3}$$

where the symbols $\{u\}$ *obey the ordinary laws of algebra, viz.,*

$$
\begin{aligned}
&a\{u\} + b\{u\} = (a + b)\{u\},\\
(3.4)\qquad &\{u\}\{v\} = \{v\}\{u\} = \{uv\},\\
&(a\{u\})(b\{v\}) = ab\{uv\}.
\end{aligned}
$$

PROOF. From the conditions satisfied by the block $B^{(1)}$ of N_1 we find that the matrix $N_1'N_1$ contains a row of the form

$$(3.5)\qquad \rho_1 = (0, 0, \cdots, 0, \quad 1, 1, \cdots, 1, \cdots, \quad k_1, k_1, \cdots, k_1),$$

where the integer i is repeated $b_i^{(1)}$ times; $i = 0, 1, \cdots, k_1$. Similarly from the properties of the block $B^{(2)}$ of N_2 we find that the matrix $N_2'N_2$ contains a row of the form

$$(3.6)\qquad \rho_2 = (0, 0, \cdots, 0, \quad 1, 1, \cdots, 1, \cdots, \quad k_2, k_2, \cdots, k_2),$$

where the integer j occurs $b_j^{(2)}$ times; $j = 0, 1, \cdots, k_2$. The matrix $N'N$ for the Kronecker product $N = N_1 \times N_2$ will clearly contain a row $\rho = \rho_1 \times \rho_2$.

Now pick out the integer 0 in ρ. It arises b_2 times when each of the $b_0^{(1)}$ zeros in ρ_1 is the coefficient of ρ_2 in ρ, and also b_1 times when each of the $b_0^{(2)}$ zeros in ρ_2 multiplies the elements of ρ_1. But in this enumeration of zeros, the multiplication of zeros of ρ_1 and ρ_2 has been counted twice, so that actually the number of zeros in ρ is $b_0^{(0)} = b_0^{(2)}b_1 + b_0^{(1)}b_2 - b_0^{(1)}b_0^{(2)}$, which is exactly the coefficent of $\{0\}$ in (3.3) when expanded according to the properties (3.4).

Similarly, the integer 1 occurs only in those places where one of the $b_1^{(1)}$ 1's of ρ_1 multiplies one of the $b_1^{(2)}$ 1's in ρ_2. Hence we must have $b_1^{(0)} = b_1^{(1)} \cdot b_1^{(2)}$ which is exactly the coefficient of $\{1\}$ in the expression (3.3) when expanded according to the properties (3.4).

In the same way, it is easy to verify that the integer u occurs in ρ $b_u^{(0)}$ times where $b_u^{(0)}$ is the coefficient of $\{u\}$ in (3.3). This proves the theorem.

From the block structures of affine resolvable BIB designs [2] and symmetrical BIB designs [1], we can easily deduce the following corollaries of Theorem 3.2.

COROLLARY 3.2.1. *If a* PBIB *design with three associate classes and with parameters* (2.5) *is the Kronecker product of the affine resolvable* BIB *design with parameters*

$$
(3.7)\qquad
\begin{aligned}
&v_1 = nk_1 = n^2\{(n-1)t + 1\},\\
&\qquad b_1 = nr_1 = n\{n^2t + n + 1\}, \qquad \lambda_1 = nt + 1,
\end{aligned}
$$

and the symmetrical BIB *design with parameters*

$$(3.8)\qquad v_2 = b_2, \qquad r_2 = k_2, \qquad \lambda_2,$$

then with respect to any block B in it, the other blocks fall into four groups (α), (β), (γ), *and* (δ) *such that the group* (α) *contains* $n_1' = b_2 - 1 = v_2 - 1$ *blocks each having* $k_1\lambda_2$ *treatments in common with B, the group* (β) *contains* $b_1 - n$ *blocks each having* mr_2 *treatments in common with B, the group* (γ) *contains* $n_1'(b_1 - n)$

blocks each having $m\lambda_2$ treatments in common with B, and the group (δ) contains $b_2(n-1)$ blocks each having zero treatments in common with B, where

$$m = (k_1)^2/v_1 = (n-1)t + 1.$$

The groups (α), (β), (γ), (δ) are all distinct if $n\lambda_2 \neq k_2$.

COROLLARY 3.2.2. *If a* PBIB *design with three associate classes and with parameters* (2.5) *is the Kronecker product of the two affine resolvable* BIB *designs with parameters.*

$$(3.9)\qquad \begin{aligned} v_1 &= n_1k_1 = n_1^2\{(n_1-1)t_1+1\}, \\ b_1 &= n_1r_1 = n_1\{n_1^2t_1+n_1+1\}, \qquad \lambda_1 = n_1t_1+1 \end{aligned}$$

and

$$(3.10)\qquad \begin{aligned} v_2 &= n_2k_2 = n_2^2\{(n_2-1)t_2+1\}, \\ b_2 &= n_2r_2 = n_2\{n_2^2t_2+n_2+1\}, \qquad \lambda_2 = n_2t_2+1, \end{aligned}$$

then with respect to any block B in it, the other blocks fall into four groups (α), (β), (γ) and (δ), such that the group (α) contains $b_2 - n_2$ blocks each having m_2k_1 treatments in common with B, the group (β) contains $b_1 - n_1$ blocks each having m_1k_2 treatments in common with B, the group (γ) contains $(b_1-n_1)(b_2-n_2)$ blocks each having m_1m_2 treatments in common with B, and the group (δ) contains

$$b_1(n_2-1) + b_2(n_1-1) - (n_1-1)(n_2-1)$$

blocks each having zero treatments in common with B, where

$$m_1 = (k_1)^2/v_1 = (n_1-1)t_1 + 1 \quad \text{and} \quad m_2 = (k_2)^2/v_2 = (n_2-1)t_2 + 1.$$

The groups (α), (β), (γ), (δ) are all distinct if $n_1 \neq n_2$.

4. Concluding remarks.

(i) A similar analysis can be carried out for the PBIB designs with two associate classes which are Kronecker product of BIB designs (cf. 2B above).

(ii) It is easy to see that a PBIB design which is the Kronecker product of a resolvable BIB design and another BIB design is also resolvable.

(iii) It is interesting to note clearly the connection between corollaries to Theorem 3.2 on the one hand and Theorem 2.1 on the inversion of designs on the other. For example, from Corollary 3.2.1 one may gather the false impression that the Kronecker product of an affine resolvable BIB design and a symmetrical BIB design would lead on inversion to a PBIB design with four associate classes. Remembering, however, that an affine resolvable BIB design gives on inversion a PBIB design with two associate classes and that a symmetrical BIB design is self-dual, we find from Theorem 2.1 and Theorem 4.2 of [7], that the dual of the Kronecker product under consideration is, in fact, a PBIB design with five associate classes all of which are distinct. This apparent contradiction is resolved if we observe that the number of distinct associate classes in a PBIB design depends not only on its λ parameters but also on the matrices

(p^i_{ju}) of its secondary parameters, the exact relation being given in Lemma 4.1 of [7], whereas for finding relations among the blocks of the inverted design we are concerned only with the number of different λ parameters of the PBIB design. Thus in the example under discussion, the PBIB design with five distinct associate classes has two of its λ parameters equal to zero, and therefore there are only four different λ parameters which determine the four types of relations among the blocks of the inverted design.

Similar remarks apply to the PBIB designs obtained in 2B and their duals.

Further work of this type applicable to PBIB designs in general is under progress and the author hopes to publish a separate paper dealing with it.

Acknowledgment. I wish to express my sincere thanks to Professor M. C. Chakrabarti for his kind interest in this work. I am also indebted to the referee for his helpful suggestions.

REFERENCES

[1] R. A. Fisher, "An examination of the different possible solutions of a problem in incomplete blocks," *Ann. Eugenics*, Vol. 10 (1940), pp. 52–75.

[2] R. C. Bose, "A note on resolvability of balanced incomplete block designs," *Sankhyā*, Vol. 6 (1942), pp. 105–110.

[3] W. S. Connor, Jr., "On the structure of balanced incomplete block designs," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 57–71.

[4] R. C. Bose and W. S. Connor, "Combinatorial properties of group divisible incomplete block designs," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 367–383.

[5] Purnendu Mohan Roy, "Inversion of incomplete block designs," *Bull. Cal. Math. Soc.*, Vol. 46 (1954), pp. 47–58.

[6] S. S. Shrikhande, "On the dual of some balanced incomplete block designs," *Biometrics*, Vol. 8 (1952), pp. 66–72.

[7] Manohar Narhar Vartak, "On an application of Kronecker product of matrices to statistical designs," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 420–438.

THE DUAL OF A BALANCED INCOMPLETE BLOCK DESIGN

P. V. RAO

College of Science, Nagpur, India

1. Summary. Shrikhande [9] and Roy [7] have shown that certain Balanced Incomplete Block Designs (BIBDs) can be dualised to give Partially Balanced Incomplete Block Designs (PBIBDs) with exactly two associate classes. Roy and Laha [8] have obtained a necessary and sufficient condition for the dual of a BIBD to be a PBIBD with two associate classes. In this paper, a general result regarding the dual of a BIBD is established and the results of Shrikhande and Roy are obtained as particular cases. An illustration to show the use of the result when the dual is not a 2-associate PBIBD is also given.

2. Two Lemmas connecting the parameters of a BIBD. For the definition of a BIBD the reader may refer to Kempthorne [4]. The following two lemmas will be stated without proof. Lemma 2.1 is due to Connor [2], while Lemma 2.2 is due to Hussain [3].

LEMMA 2.1: *If l_{ij} is the number of treatments in common with the ith and the jth blocks of a* BIBD *with parameters v^*, b^*, r^*, k^*, λ^*; the following inequalities hold*:

$$[2\lambda^*k^* + r^*(r^* - \lambda^* - k^*)]/r^* \geqq l_{ij} \geqq -(r^* - \lambda^* - k^*). \tag{2.1}$$

LEMMA 2.2: *If n_u denotes the number of blocks having $u - 1$ treatments in common with a chosen initial block of a* BIBD *with parameters v^*, b^*, r^*, k^*, λ^*, and t is the largest integer contained in $[2\lambda^*k^* + r^*(r^* - \lambda^* - k^*)]/r^*$, such that $t < k + 1$, the following equalities hold*:

$$\sum_{u=1}^{t+1} n_u = b^* - 1, \tag{2.2}$$

$$\sum_{u=1}^{t+1} (u - 1)n_u = k^*(r^* - 1), \tag{2.3}$$

$$\sum_{u=1}^{t+1} (u - 1)(u - 2)n_u = k^*(k^* - 1)(\lambda^* - 1). \tag{2.4}$$

Note that if (2.2), (2.3) and (2.4) admit a unique nonnegative integral solution, then, corresponding to each block of the design, the remaining $b^* - 1$ blocks may be divided into $t + 1 = m$ groups such that a block in the uth group has exactly $u - 1 = \lambda_u$ ($u = 1, 2, \cdots m$) treatments in common with the chosen initial block, there being exactly n_u blocks in the uth group.

3. The definition of a PBIBD. An incomplete block design is said to be a PBIBD if it satisfies the following conditions:

(3.1) There are v treatments divided into b blocks of k plots each, different treatments being applied to the plots in the same block.

Received March 20, 1959; revised January 22, 1960.

(3.2) Each treatment occurs in exactly r blocks.

(3.3) There can be established an association relationship between any two treatments satisfying the following conditions:

(3.3a) Two treatments are either 1st, 2nd, $\cdots$ mth associates.

(3.3b) Each treatment has exactly n_u uth associates ($u = 1, \cdots m$).

(3.3c) Given any two treatments which are kth associates, the number of treatments which are the uth associates of the first and u'th associates of the second is $P^k_{uu'}$. Also, $P^k_{uu'} = P^k_{u'u}$.

(3.4) Two treatments which are uth associates will occur together in exactly $\lambda_u(u = 1, 2, \cdots, m)$ blocks.

For the necessary conditions satisfied by the parameters of a PBIBD the reader is referred to Bose and Nair [1] and Nair and Rao [6].

4. The dual of a design. Let $B_1, B_2, \cdots, B_{b^*}$ and $T_1, T_2, \cdots, T_{v^*}$ denote the blocks and treatments of a given design, D^*, in which $v^*(=b)$ treatments are arranged in $b^*(=v)$ blocks of $k^*(=r)$ plots each such that every treatment is replicated $r^*(=k)$ times. Let D be a new design with v treatments and b blocks constructed by placing the treatment numbered i in block numbered j of D, if in D^* the block B_i contains the treatment T_j. The designs D^* and D are said to be the duals of each other. Evidently, in D each block contains k plots and each treatment is replicated r times. Further, if $N^* = (n_{ij})$, $(i = 1, 2, \cdots, v^*; j = 1, 2, \cdots, b^*)$, where n_{ij} denotes the number of times the ith treatment occurs in the jth block, is the incidence matrix of D^*, the incidence matrix of D is $(N^*)'$, where $(N^*)'$ is the transpose of N^*. Also the element in the ith row and the jth column of the $v^* \times v^*$ matrix $(N^*)'N^*$ will be equal to the number of blocks in the dual design D in which the ith and the jth treatments occur together.

5. The dual of a BIBD. Consider a BIBD with parameters $v^*(=b)$, $b^*(=v)$, $r^*(=k)$, $k^*(=r)$, λ^*. Let $N^* = (n_{ij})$ be the incidence matrix. We have, by the well known properties of a BIBD,

$$N^*(N^*)' = \lambda^* E_{v^*} + (r^* - \lambda^*)I_{v^*}, \tag{5.1}$$

where E_{v^*} is a $v^* \times v^*$ matrix with all elements unity and I_{v^*} is a $v^* \times v^*$ identity matrix. Also,

$$(N^*)'N^* = \left(\sum_{i=1}^{v^*} n_{ij}n_{ij'}\right) = (\lambda_{jj'}), \tag{5.2}$$

where, as already observed in the previous section, $\lambda_{jj'}$ is the number of treatments common to the jth and the j'th blocks of the original BIBD, which is also equal to the number of blocks of the dual design in which the jth and the j'th treatments occur together. Thus, in the dual design, a pair of treatments can occur together in at most t blocks, where t is defined as in Lemma 2.2. Further, if the equations (2.2), (2.3) and (2.4) admit a unique integral non-negative solution, in the dual design, corresponding to each treatment, the remaining

$v - 1$ treatments can be divided into $t + 1 = m$ groups, such that a treatment in the uth group will occur in exactly $\lambda_u = u - 1$ $(u = 1, 2, \cdots, m)$ blocks with the initial treatment, and, there will be exactly n_u treatments in the uth class. At this point, it may be noted that we do not exclude the possibility of some of the n_u's being zero, in which case the exact number of classes will be less than m. In fact, the total number of groups will be exactly equal to the total number of non-null n_u's.

We now proceed to investigate the conditions under which the dual will be a PBIBD. Evidently, if the equations (2.2), (2.3) and (2.4) admit a unique integral non-negative solution, then the conditions (3.1), (3.2), (3.3a), (3.3b) and (3.4) are satisfied by the dual design. Hence it remains to see when (3.3c) will also be satisfied.

Define m $v \times v$ matrices $B_u (u = 1, 2, \cdots, m)$ as

$$(5.3) \qquad B_u = (b^u_{jj'}) \qquad j, j' = 1, 2, \cdots, v;$$

where $b^u_{jj} = 0$ for all j, and $b^u_{jj'} = 1$ if $\lambda_{jj'} = \lambda_u$ and 0 otherwise, for all $j \neq j'$.

The matrices B_u are symmetric, independent, and commutative with respect to multiplication. It is also clear that

$$(5.4) \qquad \sum_{s=1}^{v} b^u_{is} b^{u'}_{sj} = \sum_{s=1}^{v} b^u_{si} b^{u'}_{js} = C^{uu'}_{ij},$$

which is the number of treatments common to the uth and u'th groups of treatments with respect to the treatments numbered i and j in the dual design if $i \neq j$. It equals n_u if $i = j$ and $u = u'$, and it equals zero if $i = j$ and $u \neq u'$.

Now consider any block, B_i, of the original BIBD. There will be n_u blocks in the design that have exactly λ_u treatments in common with B_i. Of these n_u blocks, $C^{uu'}_{ij}$ blocks will have $\lambda_{u'}$ treatments in common with the block B_j Hence

(5.5) $\sum_{u'=1}^{m} C^{uu'}_{ij} = n_u$ if the blocks B_i and B_j do not have λ_u treatments in common,

$= n_u - 1$ otherwise.

Now using (5.2) and (5.3), and observing that $\lambda_{jj} = k^*$, we have,

$$(5.6) \qquad (N^*)'N^* = k^* I_{b^*} + \sum_{u=1}^{m} \lambda_u B_u,$$

and hence,

$$\begin{aligned}(5.7) \qquad [(N^*)'N^*][(N^*)'N^*] &= (N^*)'[N^*(N^*)']N^* \\ &= (N^*)'[\lambda^* E_{v^*} + (r^* - \lambda^*) I_{b^*}]N^* \\ &= \lambda^*(N^*)'E_{v^*}N^* + (r^* - \lambda^*)(N^*)'N^*.\end{aligned}$$

As N^* is the incidence matrix of a BIBD it is easy to verify that

$$(5.8) \qquad (N^*)'E_{v^*}N^* = (k^*)^2 E_{b^*},$$

and that the left hand side of (5.7) can also be expressed as

$$(N^*)'N^*\left[k^*I_{b^*} + \sum_{u=1}^{m} \lambda_u B_u\right]. \tag{5.9}$$

Hence, using (5.8) and (5.9) and noting that $\lambda_1 = 0$, we get from (5.7),

$$k^*(N^*)'N^* + (N^*)'N^*\left(\sum_{u=2}^{m} \lambda_u B_u\right) = \lambda^*(k^*)^2 E_{b^*} + (r^* - \lambda^*)(N^*)'N^*.$$

Hence, from (5.6),

$$\lambda^*(k^*)^2 E_{b^*} - k^*(k^* - r^* - \lambda^*)I_{b^*} = (2k^* - r^* + \lambda^*)\sum_{u=2}^{m} \lambda_u B_u + \left[\sum_{u=2}^{m} \lambda_u B_u\right]^2.$$

Hence

$$\lambda^*(k^*)E_{b^*} - k^*(r^* - k^* - \lambda^*)I_b{}^* = (2k^* - r^* + \lambda^*)\sum_u \lambda_u B_u + \sum_{u,u'}\sum \lambda_u\lambda_{u'}B_uB_{u'}. \tag{5.10}$$

Comparing the (ij)th non-diagonal terms on both sides of (5.10),

$$\sum_{u,u'}\sum \lambda_u\lambda_{u'}\sum_s b_{is}^u b_{sj}^{u'} = \lambda^*(k^*)^2 - (2k^* - r^* + \lambda^*)\sum_u \lambda_u b_{ij}^u.$$

Using the notation of (5.4),

$$\sum_{u,u'}\sum \lambda_u\lambda_{u'}C_{ij}^{uu'} = \lambda^*(k^*)^2 - (2k^* - r^* + \lambda^*)\sum \lambda_u b_{ij}^u. \tag{5.11}$$

We can divide the set of $(b^*)^2$ equations (5.11) into m mutually exclusive sets such that the qth set $(q = 1, 2, \cdots, m)$ contains all the equations with $C_{ij}^{uu'}$ for $\lambda_{ij} = \lambda_q$ The coefficients in the left hand side, and the constant in the right hand side, are same for all the equations in a given set. In fact, the equations in the qth set will be obtained by giving all the values to i and j such that $\lambda_{ij} = \lambda_q$ in

$$\sum_{u,u'}\sum \lambda_u\lambda_{u'}C_{ij}^{uu'} = \lambda^*(k^*)^2 - (2k^* - r^* + \lambda^*)\lambda_q. \tag{5.12}$$

Thus it is clear that the values of $C_{ij}^{uu'}$ depend only on λ_u, $\lambda_{u'}$ and λ_{ij}. Hence, by writing $C_{ij}^{uu'} = P_{uu'}^q$ if $\lambda_{ij} = \lambda_q$, the equations (5.6) and (5.12) may be rewritten as

$$\sum_{u'=1}^{m} P_{uu'}^q = n_u \quad \text{if } u \neq q, \qquad = n_u - 1 \quad \text{if } u = q; \tag{5.13}$$

and

$$\sum_{u,u'}\sum \lambda_u\lambda_{u'}P_{uu'}^q = \lambda^*(k^*)^2 - (2k^* - r^* + \lambda^*)\lambda_q, \quad q = 1, 2, \cdots, m. \tag{5.14}$$

Hence, if (5.14) has a unique integral non-negative solution, it follows from (5.4) and (5.13) that the number of treatments common to the uth group and u'th

group of two treatments is the same for all treatment pairs which belong to the qth group with respect to each other. This number is equal to $P^q_{uu'}$ with $P^q_{uu'} = P^q_{u'u}$. Thus we have proved Theorem 5.1.

THEOREM 5.1: *The dual of a* BIBD *with parameters* $v^*(=b)$, $b^*(=v)$, $r^*(=k)$, $k^*(=r)$, λ^* *is a* PBIBD *with parameters* $v, b, r, k; \lambda_1, \lambda_2, \cdots, \lambda_m; n_1, n_2, \cdots, n_m; P^q_{uu'}(u, u', q = 1, 2, \cdots, m)$, *where* $m = t + 1$ *is defined as in Lemma* 2.2, *provided the equations* (2.2), (2.3), (2.4) *and* (5.14) *admit unique integral non-negative solution subject to the conditions* (5.13).

6. Shrikhande's two theorems as particular cases of the Theorem 5.1.

(6.1) *The case* $\lambda^* = 1$. Consider a BIBD with parameters $v^*(=b)$, $b^*(=v)$, $r^*(=k)$, $k^*(=r)$, $\lambda^* = 1$. In this case we have $t = 1$ and the equations (2.2), (2.3) and (2.4) reduce to $n_1 + n_2 = b^* - 1$ and $n_2 = k^*(r^* - 1)$, giving the unique non-negative solution

$$n_1 = (v - 1) - r(k - 1),$$
$$n_2 = r(k - 1).$$

Noting that $\lambda_1 = 0$ and $\lambda_2 = 1$, we can solve the equations (5.14) uniquely to get the solution $P^1_{22} = r^2$, $P^2_{22} = r^2 - 2r + k - 1 = (r - 1)^2 + (k - 2)$. The other parameters can be easily obtained by using condition (5.13).

Thus we have proved Shrikhande's [9] Theorem 1 that the dual of a BIBD with parameters $v^* = rk - k + 1$, $b^* = k(rk - k + 1)/r$, $r^* = k$, $k^* = r$, $\lambda^* = 1$ is a PBIBD with parameters $v = k(rk - k + 1)/r$, $b = rk - k + 1$, $r = r$, $k = k$, $\lambda_1 = 0$, $\lambda_2 = 1$; $n_1 = r(k - 1)$, $n_2 = (k - r)(r - 1)(k - 1)/r$;

$$p^1_{uu'} = \begin{bmatrix} (k - r)^2 + 2(r - 1) - k(k - 1)/r & r(k - r - 1) \\ r(k - r - 1) & r^2 \end{bmatrix}$$

$$p^2_{uu'} = \begin{bmatrix} (r - 1)(k - r)(k - r - 1)/r & (r - 1)(k - r) \\ (r - 1)(k - r) & (k - 2) + (r - 1)^2 \end{bmatrix}.$$

(6.2) *The case* $\lambda^* = 2$. It can be easily seen that, if we exclude the solutions in which the same block is repeated, for all designs with $\lambda^* = 2$ and $r \leqq 10$, we must have $t = 2$. In this case the equations (2.2), (2.3) and (2.4) will have the unique solution given by

$$n_1 = (b^* - 1) - k^*(r^* - k^*) - k^*(k^* - 1)/2,$$
$$n_2 = k^*(r^* - k^*),$$
$$n_3 = k^*(k^* - 1)/2.$$

But, in general, equations (5.14) will not have a unique solution. However, if we consider the particular case $n_1 = 0$, i.e. when $r^* = k^* + 2$, the equations (5.14), when $q = 3$, reduce to $P^3_{22} + 4(P^3_{23} + P^3_{33}) = 2k^*(k^* - 1)$. Hence, using (5.13), we get, $P^3_{22} = 2k^*(k^* - 1) - 4(n_3 - 1) = 4$. Similarly, the other parameters may be found. Hence we have proved Theorem 3 of Shrikhande [6]

that the dual of a BIBD with parameters

$$v^* = \binom{k-1}{2}, \quad b^* = \binom{k}{2}, \quad r^* = k, \quad k^* = k-2, \quad \lambda^* = 2,$$

is a PBIBD with parameters

$$v = \binom{k}{2}, \quad b = \binom{k-2}{2}, \quad r = k-2, \quad k = k;$$

$$\lambda_1 = 1, \quad \lambda_2 = 2; \quad n_1 = 2(k-2), \quad n_2 = \binom{k-2}{2};$$

$$P^1_{uu'} = \begin{bmatrix} k-2 & k-3 \\ k-3 & \binom{k-3}{2} \end{bmatrix}; \quad P^2_{uu'} = \begin{bmatrix} 4 & 2(k-4) \\ 2(k-4) & \binom{k-4}{2} \end{bmatrix}.$$

Roy's [7] Theorem 3, regarding the dual of an affine resolvable BIBD, can be proved in a similar way by using Theorem 5.1 of this paper.

7. Application of Theorem 5.1 when the solution of the equations is not unique. When the solution of the equations (2.2), (2.3), (2.4) and (5.14) is not unique, Theorem 5.1 will not give complete information about the dual. However, if the structure of the original BIBD is known, Theorem 5.1 can be used to simplify the investigation about the properties of the dual. As an illustration, we consider the dual of a BIBD with parameters $v^* = 16$, $b^* = 24$, $r^* = 9$, $k^* = 6$, $\lambda^* = 3$. A plan of this design is given by Mann [5]. He constructed it by the process of residuation from the symmetric BIBD with parameters $v^* = b^* = 25$, $r^* = k^* = 9$, $\lambda^* = 3$. We shall denote Mann's design by D^*.

Since any two blocks of a symmetric BIBD must have λ^* treatments in common, any two blocks of the design D^* cannot have more than three treatments in common. Hence we must have $n_5 = n_6 = n_7 = 0$. Thus the equations (2.2), (2.3), and (2.4) can be written as

$$n_1 = 5 - n_4,$$

$$n_2 = 3\,(n_4 - 4),$$

$$n_3 = 3\,(10 - n_4).$$

From inspection of Mann's plan we can see that no two blocks of the design D^* have exactly one treatment in common. This gives the unique solution, $n_1 = 1$, $n_2 = 0$, $n_3 = 18$, $n_4 = 4$. For the sake of simplicity, we shall write n_1, n_2, n_3 instead of n_1, n_3, n_4 and make corresponding changes in $P^q_{uu'}$. Now, as $n_1 = 1$ and $P^1_{11} + P^1_{12} + P^1_{13} = n_1 - 1 = 0$, we must have $P^1_{11} = P^1_{12} = P^1_{13} = 0$. Therefore, the equations (5.14), when $q = 1$, may be solved uniquely to get the values of $P^1_{uu'}(u, u' = 2, 3.)$. Again, if the dual of the design D^* is a PBIBD, then P^2_{12} and P^3_{13} must both be unique and equal to $(n_1/n_2)P^1_{22}$ and $(n_1/n_3)P^1_{33}$ respectively. It can be verified that, for the design D^*, the values of P^2_{12} and P^3_{13} satisfy

these conditions and are both equal to 1. Hence, as $n_1 = 1$, it follows from (5.13) that $P^2_{11} = P^2_{13} = P^3_{11} = P^3_{12} = 0$. It is now easy to see that the equations (5.14) will have the unique solution

$$P^1_{uu'} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 18 & 0 \\ 0 & 0 & 4 \end{bmatrix},$$

$$P^2_{uu'} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 12 & 4 \\ 0 & 4 & 0 \end{bmatrix},$$

$$P^3_{uu'} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 18 & 0 \\ 1 & 0 & 2 \end{bmatrix}.$$

Hence the dual of the design D^* is a PBIBD with the parameters $v = 24$, $b = 16$, $r = 6$, $k = 9$; $\lambda_1 = 0$, $\lambda_2 = 2$, $\lambda_3 = 3$; $n_1 = 1$, $n_2 = 18$, $n_3 = 4$; $P^q_{uu'}(u, u', q = 1, 2, 3.)$.

Roy and Laha [8] have already pointed out that this PBIBD may be obtained as the dual of a BIBD. However, they have not stated how they arrived at this conclusion.

REFERENCES

[1] R. C. Bose and K. R. Nair, "Partially balanced incomplete block designs," *Sankhyā* Vol. 4 (1939), pp. 337–373.

[2] W. S. Connor, "On the structure of balanced incomplete block design," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 57–71.

[3] Q. M. Hussain, "Structure of some incomplete block designs," *Sankhyā*, Vol. 8 (1948) pp. 381–383.

[4] O. Kempthorne, *The Design and Analysis of Experiments*, John Wiley and Sons, New York, 1952.

[5] H. B. Mann, *Analysis and Design of Experiments*, Dover Pubs., New York, 1949.

[6] K. R. Nair and C. R. Rao, "A note on partially balanced incomplete block designs," *Science and Culture*, Vol. 7 (1942), p. 568.

[7] Purnendu Mohan Roy, "On the method of inversion in the construction of partially balanced incomplete block designs from the corresponding b.i.b. designs," *Sankhyā* Vol. 14 (1954), pp. 39–52.

[8] J. Roy and R. G. Laha, "Classification and analysis of linked block designs," *Sankhyā* Vol. 17 (1956), pp. 115–132.

[9] S. S. Shrikhande, "On the dual of some balanced incomplete block designs," *Biometrics* Vol. 8 (1952), pp. 66–71.

NOTES

ON THE UNBIASEDNESS OF YATES' METHOD OF ESTIMATION USING INTERBLOCK INFORMATION[1]

By Franklin A. Graybill and V. Seshadri

Oklahoma State University

In a balanced incomplete block model with blocks and errors random normal variables, Yates has shown that there are two independent unbiased estimates for any treatment contrast. These are referred to as intrablock and interblock estimators. Yates has also given a method for combining these two estimators which depends on the variances (unknown) and has shown how to estimate the variances from an analysis of variance [1]. Since this combined estimator is used quite extensively, it seems desirable to study its properties. Graybill and Weeks [2] have shown that Yates' combined estimator is based on a set of minimal sufficient statistics and have presented an estimator which is unbiased.

The purpose of this note is to show that Yates' estimator, which is based on intrablock and interblock information, is unbiased.

The model and distributional assumptions in this paper are exactly those given in [2], and the same notations are used and will not be repeated here.

In [2] it is shown that Yates' estimator (denoted by $\bar{\tau}_i$) of τ, is

$$(1)\qquad \begin{aligned} \bar{\tau}_i &= x_i + \gamma(u_i - x_i) &\quad \text{if } \hat{\sigma}_\beta^2 > 0 \\ &= x_i + \lambda t/rk(u_i - x_i) &\quad \text{if } \hat{\sigma}_\beta^2 \leqq 0 \end{aligned}$$

where

$$(2)\qquad \gamma = \frac{\dfrac{\lambda^2 t(r-\lambda)}{rk(r-1)}(U-X)'(U-X) + \dfrac{\lambda k}{(r-1)}S^{*2} + \dfrac{\lambda(k-t)}{f(r-1)}S^2}{\dfrac{\lambda^2 t(r-\lambda)}{rk(r-1)}(U-X)'(U-X) + \dfrac{\lambda k}{(r-1)}S^{*2} + \left[\dfrac{\lambda(k-t)}{f(r-1)} + \dfrac{(r-\lambda)}{f}\right]S^2}$$

and where

$$(3)\qquad \hat{\sigma}_\beta^2 = 1/t(r-1)[\lambda t(r-\lambda)/rk^2(U-X)'(U-X) + S^{*2} - (b-1)/fS^2]$$

We now define $\phi(\hat{\sigma}_\beta^2)$ such that

$$\begin{aligned} \phi(\hat{\sigma}_\beta^2) &= 0 &\quad \text{if } \hat{\sigma}_\beta^2 > 0 \\ &= 1 &\quad \text{if } \hat{\sigma}_\beta^2 \leqq 0 \end{aligned}$$

Received December 11, 1959; revised May 10, 1960.

[1] This work was sponsored in part by the Research Foundation of Oklahoma State University.

Yates' estimate can now be written as

$$(4) \quad \hat{\tau}_i = [1 - \phi(\hat{\sigma}_\beta^2)][x_i + \gamma(u_i - x_i)] + \phi(\hat{\sigma}_\beta^2)[x_i + (\lambda t/rk)(u_i - x_i)]$$

Clearly (4) is equivalent to (1). Rearranging and simplifying (4) we get

$$\hat{\tau}_i = [x_i + \gamma(u_i - x_i)] + \phi(\hat{\sigma}_\beta^2)[(\lambda t/rk) - \gamma](u_i - x_i)$$

Graybill and Weeks have shown in [2] that $E[x_i + \gamma(u_i - x_i)] = \tau_i$. Therefore in order to show that Yates' estimate is unbiased we need only show that

$$E[\phi(\hat{\sigma}_\beta^2)((\lambda t/rk) - \gamma)(u_i - x_i)] = 0$$

Let $z_i = (u_i - x_i)$ where $i = 1, 2, \cdots, t - 1$. Now $\hat{\sigma}_\beta^2$ is a function of z_i, S^{*2}, and S^2. So let

$$\hat{\sigma}_\beta^2 = g(z_1, z_2, \cdots, z_{t-1}, S^{*2}, S^2).$$

γ is also a function of z_i, S^{*2}, and S^2. Therefore, let

$$\gamma = h(z_1, z_2, \cdots, z_{t-1}, S^{*2}, S^2).$$

Denote the joint density of the $t + 1$ random variables $z_1, z_2, \cdots, z_{t-1}, S^{*2}, S^2$ by $f(z_1, z_2, \cdots, z_{t-1}, S^{*2}, S^2)$. From (2) it is clear that γ is an even function of the z_i and from (3) we see that $\hat{\sigma}_\beta^2$ is also an even function of the z_i. Therefore, $\phi(\hat{\sigma}_\beta^2)$ is an even function of z_i, $(i = 1, 2, \cdots, t - 1)$ and $\phi(\hat{\sigma}_\beta^2)[(\lambda t/rk) - \gamma]$ is also an even function of z_i. Hence $\phi(\hat{\sigma}_\beta^2)[(\lambda t/rk) - \gamma](u_i - x_i)$ is an odd function of z_i. Therefore,

$$E[\phi(\hat{\sigma}_\beta^2)((\lambda t/rk) - \gamma)(u_i - x_i)] = 0,$$

since z_i are independent normal variables with mean zero and are independent of S^2 and S^{*2}. Thus Yates' estimator, which is based on intrablock and interblock information, is unbiased.

REFERENCES

[1] Yates, F. "The recovery of interblock information in balanced incomplete block designs," *Ann. Eugenics*. Vol. 10, (1940), pp. 317–325.

[2] Graybill, F. A., and D. L. Weeks, "Combining interblock and intrablock information in balanced incomplete blocks," *Ann. Math. Stat.* Vol. 30, (1959), pp. 799–805.

ON THE BLOCK STRUCTURE OF CERTAIN PBIB DESIGNS WITH TWO ASSOCIATE CLASSES HAVING TRIANGULAR AND L_2 ASSOCIATION SCHEMES

By Damaraju Raghavarao

University of Bombay

0. Summary. The PBIB designs [2] with two associate classes are classified in [3] as 1. Group Divisible, 2. Simple, 3. Triangular, 4. Latin Square type with i

Received November 6, 1959.

constraints, and 5. Cyclic. Group Divisible designs are divided into three types [1]: 1. Singular, 2. Semi-regular, and 3. Regular. It has been proved [1] that every block of a Semi-regular Group Divisible design contains k/m treatments from each of the m groups of the association scheme. In this note we prove analogous results in the case of certain PBIB designs with triangular and L_2 association schemes.

1. On the Block Structure of certain PBIB designs with two associate classes having a triangular association scheme. A PBIB design with two associate classes is said to have a triangular association scheme [3] if the number of treatments $v = n(n - 1)/2$ and the association scheme is an array of n rows and n columns with the following properties:

(a) The positions in the principal diagonal are blank.

(b) The $n(n - 1)/2$ positions above the principal diagonal are filled by the numbers $1, 2, \cdots, n(n - 1)/2$ corresponding to the treatments.

(c) The array is symmetric about the principal diagonal.

(d) For any treatment θ, the first associates are exactly those treatments which lie in the same row and same column as θ.

It is then obvious that

(1) the number of first associates of any treatment is $n_1 = 2n - 4$, and

(2) with respect to any two treatments θ_1 and θ_2 which are first associates, the number of treatments which are first associates of both θ_1 and θ_2 is $p_{11}^1(\theta_1, \theta_2) = n - 2$.

We now prove

THEOREM 1.1. *If in a PBIB design with two associate classes having a triangular association scheme*

$$rk - v\lambda_1 = n(r - \lambda_1)/2, \tag{1.1}$$

then $2k$ is divisible by n. Further, every block of the design contains $2k/n$ treatments from each of the n rows of the association scheme.

PROOF. Let e_j^i treatments occur in the jth block from the ith row of the association scheme ($i = 1, 2, \cdots, n$; $j = 1, 2, \cdots, b$). Then we have

$$\begin{aligned} \sum_{j=1}^{b} e_j^i &= (n - 1)r, \\ \sum_{j=1}^{b} e_j^i(e_j^i - 1) &= (n - 1)(n - 2)\lambda_1, \end{aligned} \tag{1.2}$$

since each of the treatments occurs in r blocks and every pair of treatments from the same row of the association scheme occurs together in λ_1 blocks. From (1.2), we get

$$\sum_{j=1}^{b} (e_j^i)^2 = (n - 1)\{r + (n - 2)\lambda_1\}. \tag{1.3}$$

Define $e_{\cdot}^{i} = b^{-1}\sum_{j=1}^{b} e_j^i = (n-1)r/b = 2k/n$. Then

$$
\begin{aligned}
\sum_{j=1}^{b} (e_j^i - e_{\cdot}^{i})^2 &= (n-1)\{r + (n-2)\lambda_1\} - 4bk^2/n^2 \\
&= 2(n-1)[\{n(r-\lambda_1)/2\} - (rk - v\lambda_1)]/n \\
&= 0,
\end{aligned}
\tag{1.4}
$$

from (1.1). Therefore $e_1^i = e_2^i = \cdots = e_b^i = e_{\cdot}^i = 2k/n$. Since e_j^i $(i = 1, 2, \cdots, n;$ $j = 1, 2, \cdots, b)$ must be integral, $2k$ is divisible by n. This completes the proof of the theorem.

It has been proved ([4], [5], [7]) that a PBIB design with two associate classes satisfying the relations (1) and (2) has a triangular association scheme for all n except 8. Using this result and Theorem 1.1, we have

COROLLARY 1.1.1. *A necessary condition for the existence of a PBIB design with two associate classes having the parameters*

$$v = n(n-1)/2,\ b,\ r,\ k,\ \lambda_1,\ \lambda_2,\ n_1 = 2n-4,\ p_{11}^1 = n-2, \tag{1.5}$$

where $rk - v\lambda_1 = n(r-\lambda_1)/2$ *and* $n \neq 8$, *is that* $2k$ *is divisible by* n.

Now let us consider the PBIB design with parameters

$$
\begin{aligned}
&v = n(n-1)/2, \quad b = (n-1)(n-2)/2, \quad r = n-2, \\
&k = n, \quad n_1 = 2n-4, \quad n_2 = (n-2)(n-3)/2, \\
&\lambda_1 = 1, \quad \lambda_2 = 2, \quad p_{11}^1 = n-2, \quad p_{11}^2 = 4
\end{aligned}
\tag{1.6}
$$

This PBIB design has been shown to have a triangular association scheme [8]. Further, the parameters satisfy relation (1.1). Hence every block of this design contains $2k/n = 2$ treatments from each of the n rows of the association scheme.

2. On the Block Structure of certain PBIB Designs with two associate classes having a L_2 association scheme. A PBIB design is said to have a L_2 association scheme [3], if the number of treatments $v = s^2$, where s is a positive integer, and the treatments can be arranged in an $s \times s$ square such that treatments in the same row or the same column are first associates, while others are second associates. The following results are easily seen to hold in this case:

(i) The number of first associates of any treatment is $n_1 = 2s - 2$.

(ii) With respect to any two treatments θ_1 and θ_2 which are first associates, the number of treatments which are first associates of both θ_1 and θ_2 is $p_{11}^1 = s - 2$.

We now prove

THEOREM 2.1. *If, in a PBIB design with two associate classes having a* L_2 *association scheme,*

$$rk - v\lambda_1 = s(r - \lambda_1), \tag{2.1}$$

then k *is divisible by* s. *Further, every block of the design contains* k/s *treatments from each of the* s *rows (or columns) of the association scheme.*

PROOF. Let f_p^q treatments occur in the pth block from the qth row (or column)

of the association scheme ($p = 1, 2, \cdots, b$; $q = 1, 2, \cdots, s$). We then have

$$\begin{aligned} &\sum_{p=1}^{b} f_p^q = sr, \\ (2.2) \qquad &\sum_{p=1}^{b} f_p^q(f_p^q - 1) = s(s-1)\lambda_1, \end{aligned}$$

since each of the treatments occurs in r blocks and every pair of the treatments from the same row (or column) of the association scheme occurs together in λ_1 blocks.

From (2.2), we get

$$(2.3) \qquad \sum_{p=1}^{b} (f_p^q)^2 = s\{r + (s-1)\lambda_1\}.$$

Define $f_{\cdot}^q = b^{-1}\sum_{p=1}^{b} f_p^q = sr/b = k/s$. Then

$$\begin{aligned} (2.4) \qquad \sum_{p=1}^{b} (f_p^q - f_{\cdot}^q)^2 &= s\{n + (s-1)\lambda_1\} - bk^2/s^2 \\ &= s(r - \lambda_1) - (rk - v\lambda_1) = 0, \end{aligned}$$

from (2.1). Therefore $f_1^q = f_2^q = \cdots = f_p^q = f_{\cdot}^q = k/s$. Since f_p^q ($p = 1, 2, \cdots, b$; $q = 1, 2, \cdots, s$) must be integral, k is divisible by s. Thus the theorem is proved.

It has been proved that a PBIB design with two associate classes satisfying the relations (i) and (ii) has a L_2 association scheme if $s \neq 4$([6], [9]). Using this result and Theorem 2.1 we have

COROLLARY 2.1.1. *A necessary condition for the existence of a PBIB design with two associate classes having the parameters*

$$(2.5) \qquad v = s^2, b, r, k, \lambda_1, \lambda_2, \qquad n_1 = 2s - 2, \qquad p_{11}^1 = s - 2,$$

where $rk - v\lambda_1 = s(r - \lambda_1)$ and $s \neq 4$, is that k is divisible by s.

Acknowledgement. My sincere thanks are due to Professor M. C. Chakrabarti for his kind interest in this work.

REFERENCES

[1] R. C. BOSE AND W. S. CONNOR, "Combinatorial properties of group divisible incomplete block designs," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 367–383.

[2] R. C. BOSE AND K. R. NAIR, "Partially balanced incomplete block designs," *Sankhyā*, Vol. 4 (1939) pp. 337–372.

[3] R. C. BOSE AND T. SHIMAMOTO, "Classification and analysis of partially balanced designs with two associate classes," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 151–190.

[4] W. S. CONNOR, "The uniqueness of the triangular association scheme," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 262–266.

[5] A. J. HOFFMAN, "On the uniqueness of the triangular association scheme," *Ann. Math. Stat.*, Vol. 31 (1960), pp. 492–497.

[6] D. M. MESNER, "An investigation of certain combinatorial properties of partially balanced incomplete block designs and association schemes with a detailed study of designs of Latin Square and related types," Unpublished thesis, Michigan State University, 1956.

[7] S. S. Shrikhande, "On a characterization of the triangular association scheme," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 39–48.

[8] S. S. Shrikhande, "Relation between certain types of incomplete block designs," Institute of Statistics, University of North Carolina, *Mimeograph Series No. 207* (1958).

[9] S. S. Shrikhande, "The uniqueness of the L_2 association scheme," *Ann. Math. Stat.* Vol. 30 (1959), pp. 781–798.

OPTIMALITY CRITERIA FOR INCOMPLETE BLOCK DESIGNS[1]

By K. R. Shah[2]

Forest Research Institute, Dehra Dun, India

1. Introduction and Summary. Several optimality criteria have been suggested for the efficiency of incomplete block designs. This note surveys these criteria, extends certain results and puts forward a new and simpler criterion.

2. Existing Criteria. Important aims in experimental design are to estimate the effects of treatment comparisons with maximum precision for a given total number of experimental units, or total cost, and to perform a test of the null hypothesis. These two considerations lead us to different criteria for choosing from among the designs.

Consider the class of incomplete block designs, D_{vbk}, for fixed values of v, k and $b(v > k)$, where v treatments are arranged in b blocks of k plots each, and each treatment is replicated r times. In the usual notation, (see for example, Kempthorne [2]) intra-block estimates of treatment effects are given by

$$\mathbf{C}\hat{\mathbf{t}} = \mathbf{Q}, \tag{2.1}$$

where $\mathbf{C} = r\mathbf{I} - \mathbf{NN}'/k$, $\mathbf{N}$ being the incidence matrix of the design. We consider only connected designs, so that the rank of $\mathbf{C}$ is $v - 1$. Let $\lambda_1, \lambda_2, \cdots, \lambda_v$, be the $v - 1$ non-zero latent roots of $\mathbf{C}$. It is proved in [2] that the average variance of all elementary treatment contrasts is proportional to $\sum \lambda_i^{-1}$. Let $\mathbf{P}_i'\mathbf{t}(i = 1, 2, \cdots, v - 1)$ be any complete set of $v - 1$ orthogonal normalised contrasts. Set

$$\mathbf{P} = [\mathbf{P}_1, \mathbf{P}_2, \cdots, \mathbf{P}_{v-1}], \qquad \mathbf{P}'\mathbf{t} = \boldsymbol{\rho}, \qquad \boldsymbol{\rho} = \{\rho_1, \cdots, \rho_{v-1}\}.$$

It can be shown that $\mathbf{P}'\mathbf{CP}$ is a non-singular matrix with latent roots $\lambda_1, \cdots, \lambda_{v-1}$, and that (2.1) leads to

$$\mathbf{P}'\mathbf{CP}\hat{\boldsymbol{\rho}} = \mathbf{P}'\mathbf{Q} \quad \text{or} \quad \hat{\boldsymbol{\rho}} = (\mathbf{P}'\mathbf{CP})^{-1}\mathbf{P}'\mathbf{Q}. \tag{2.2}$$

Let us denote the dispersion matrix of $\mathbf{x}$ by $V(\mathbf{x})$. Now $V(\mathbf{Q}) = \mathbf{C}\cdot\sigma^2$, which

Received June 8, 1959.

[1] This work was supported by a senior research training scholarship of the Government of India.

[2] Present address: Indian Statistical Institute, Calcutta, India.

gives $V(\hat{\varrho}) = (\mathbf{P}'\mathbf{C}\mathbf{P})^{-1}\cdot\sigma^2$. Hence the generalised variance of $\hat{\varrho}$ is given by

$$(2.3)\qquad |V(\hat{\varrho})| = |(\mathbf{P}'\mathbf{C}\mathbf{P})^{-1}|\cdot\sigma^2 = \sigma^2\prod_{i=1}^{v-1}\lambda_i^{-1}.$$

The usual null hypothesis H_0 is $t_1 = t_2 = \cdot = t_v$, which is equivalent to $\rho_1 = \rho_2 = \cdot = \rho_{v-1} = 0$. The sum of squares for testing H_0 is $\mathbf{t}'\mathbf{Q}$, which can be shown to be equal to $\varrho'\mathbf{P}'\mathbf{C}\mathbf{P}\varrho$. Hence the power of the F test is a monotonically increasing function of $\beta = \varrho'\mathbf{P}'\mathbf{C}\mathbf{P}\varrho/\sigma^2$.

The efficiency criteria considered so far by various authors are as follows:

(A) If we wish to minimise the average variance of all elementary treatment contrasts, we should minimise $\sum\lambda_i^{-1}$, [2], [4].

(B) Wald [6] argues that it is not possible to maximise power for all values of ϱ. Hence we should maximise β for fixed values of $\varrho'\varrho/\sigma^2$. It is reasonable to maximise the minimum of β subject to $\varrho'\varrho/\sigma^2 =$ constant. This leads to maximising $\lambda_{\min}$, [1], [6].

(C) Wald [6] further argues that from certain mathematical considerations it would be simpler to minimise $\prod_{i=1}^{v-1}\lambda_i^{-1}$. This minimises the generalised variance. Also, as Nandi [5] has pointed out, this has the desirable effect of minimising the volume of equi-power ellipsoid given by $\varrho'\mathbf{P}'\mathbf{C}\mathbf{P}\varrho/\sigma^2 =$ const. In a sense this minimises the range of ϱ subject to constant power. It should also be noted that the design which minimises $\prod\lambda_i^{-1}$ gives certain optimum properties for the usual F test associated with it, [3].

It is easy to see that the optima for all the criteria are reached when the λ's are all equal. Hence, when a balanced incomplete block design (BIB) exists in the class D_{vbk}, it is the most efficient design in that class [4].

In [2] and [4] only the equi-replicate designs are considered. But the results follow from the roots of $\mathbf{C}$, and the only condition used in [4] is $\sum\lambda_i =$ constant. Hence the results in [2] and Section 2 of [4] are valid also for the case of unequal number of replications. The extension of these results is not of mere academic interest; there are important classes of designs, such as inter and intra-group block designs and reinforced incomplete block designs, where the number of replications are usually unequal.

Since efficiency should relate to the manner of utilization of the resources, in framing an efficiency criterion, it seems natural to take into account the amount of experimental material used. This would enable us to compare designs with different sizes. Hence, we consider the class of designs, D_{vk}, for fixed values of v and $k(v > k)$, where v treatments are arranged in blocks of k plots each. Denote by r_i and R, the number of replications for the ith treatment and the average number of replications respectively. Since $\sum\lambda_i = \text{Trace } \mathbf{C} = (k-1)\sum r_i/k = (k-1)vR/k$; it is linearly related to the total number of plots.

The efficiency criteria, analogous to those in (A), (B) and (C) would be

$$(2.4)\qquad E_1 = (v-1)/R\sum\lambda_i^{-1},\quad E_2 = \lambda_{\min}/R,\quad E_3 = 1/R(\prod\lambda_i)^{1/v-1}.$$

Now for fixed R, the theoretical maxima of E_1, E_2, E_3 are attained when $\lambda_1 = \lambda_2 = \cdots = \lambda_{v-1}$. Since this maximising solution is independent of R, it is

also the unconditional maximising solution. Now in the class of designs D_{vk}, a BIB design always exists. Hence, judged by any of the three critreia, within the class of designs D_{vk} any of the BIB designs is the most efficient. It can be easily seen that for the BIB design $E_1 = E_2 = E_3 = (1 - 1/k)/(1 - 1/v)$. And as is to be expected, each one of them increases with k. In the limit when $k = v$, i.e., for randomised complete block designs, $E_1 = E_2 = E_3 = 1$.

3. A New Criterion. The above three criteria are based on different considerations and need not necessarily agree in comparing two given designs. Which criterion should be adopted depends upon our aim in conducting the experiment. But most often we shall be interested in both the interval estimation of treatment effects and in the test of the null hypothesis.

It should be noted that, in the limit when optimality is reached, all the three criteria lead to the same result, *viz.*, the λ's should be all equal. In fact for the first and the third certeria, we are concerned with the geometric and the harmonic means subject to the arithmetic mean being constant. When the experiment is symmetrical, i.e., the λ's are all equal, the three means coincide. This suggests the use of $\sum (\lambda_i - \bar{\lambda})^2/(v - 1)$ with $\sum \lambda_i =$ const., as a criterion for optimality, i.e. among designs of given size, we should make $\sum \lambda_i^2$ as small as possible, subject to existence of a design. To eliminate the effect of the size of the design we define

$$(3.1) \quad E_4 = \bar{\lambda}^2/[R (\textstyle\sum \lambda_i^2/(v - 1))^{\frac{1}{2}}] = (v - 1)^{-\frac{3}{2}} (\textstyle\sum \lambda_i)^2/[R (\textstyle\sum \lambda_i^2)^{\frac{1}{2}}].$$

When the design is balanced, $E_4 = (1 - 1/k)/(1 - 1/v)$, and hence the efficiency of a BIB increases with k increasing, reaching unity when $k = v$. Nevertheless, the criterion is suggested only for comparisons of different designs within the class D_{vk}, with v and k fixed.

Though this criterion does not agree exactly with any of the three criteria given above, it will tend to be as good as any of them. In any case, we are not able to satisfy all the three criteria simultaneously. Smaller values of $\sum \lambda_i^2$ will tend to give smaller values of $\sum \lambda_i^{-1}$ and $\prod \lambda_i^{-1}$, though this does not hold exactly in all cases. Though the contours of equal efficiency (in the space of the λ's) are not identical with those for the other three criteria (which themselves are not identical), our criterion will be quite useful. For the points on the line given by $\lambda_1 = \lambda_2 = \cdots = \lambda_{v-1}$ all give the same result and for the class of designs with higher efficiency, i.e., for λ's not too widely spread, they will be more or less equal. This is the region where our criterion will be quite effective. As shown below this criterion has the advantages of simplicity and practical usefulness.

We can express $\mathbf{C}$ as $\sum \lambda_i \mathbf{L}_i \mathbf{L}_i'$, where $\mathbf{L}_i$ is the canonical vector corresponding to λ_i. This immediately gives $\mathbf{C}^2 = (\sum \lambda_i \mathbf{L}_i \mathbf{L}_i')(\sum \lambda_i \mathbf{L}_i \mathbf{L}_i') = \sum \lambda_i^2 \mathbf{L}_i \mathbf{L}_i^2$. Hence Trace $\mathbf{C}^2 = \sum \lambda_i^2$, but Trace $\mathbf{C}^2 = \sum_i \sum_j c_{ij}^2$, and therefore $\sum \lambda_i^2 = \sum_i \sum_j c_{ij}^2$. Hence,

$$E_4 = (v - 1)^{-\frac{1}{2}}((k - 1)/k)^2 v^2 R/\textstyle\sum_i \sum_j c_{ij}^2.$$

A further simplification can be had for PBIB and circulant designs, where $\sum_j c_{ij}^2$ is the same for all i.

For the other three criteria, elegant expressions are seldom available. Since E_4 follows directly from the **C** matrix it is easiest to compute; we do not have to solve the normal equations or evaluate the λ's.

Acknowledgment. My sincere thanks are due to Dr. K. R. Nair for his guidance in writing this paper.

REFERENCES

[1] Sylvain Ehrenfeld, "On the efficiency of experimental designs," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 247–255.

[2] Oscar Kempthorne, "The efficiency factor of an incomplete block design," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 846–849.

[3] J. Kiefer, "On the non-randomised optimality and randomised non-optimality of symmetrix design," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 675–699.

[4] A. M. Kshirsagar, "A note on incomplete block designs," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 907–910.

[5] H. K. Nandi, "On the efficiency of experimental designs," *Calcutta Stat. Assoc. Bull.*, Vol. 3 (1950), pp. 167–171.

[6] Abraham Wald, "On the efficient design of statistical investigations," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 134–140.

ON THE COMPLETENESS OF ORDER STATISTICS[1]

By C. B. Bell[2]

University of California, La Jolla, and San Diego State College

AND

David Blackwell and Leo Breiman

University of California, Berkeley

1. Introduction and summary. Let $X_1, X_2, \cdots, X_n$ be a sample of a one-dimensional random variable X; let the order statistic $T(X_1, X_2, \cdots, X_n)$ be defined in such a manner that $T(x_1, x_2, \cdots, x_n) = (x^{(1)}, x^{(2)}, \cdots, x^{(n)})$ where $x^{(1)} \leqq x^{(2)} \leqq \cdots \leqq x^{(n)}$ denote the ordered x's; and let Ω be a class of one-dimensional *cpf*'s, i.e., cumulative probability functions.

The order statistic, T, is said to be a complete statistic with respect to the class, $\{P^{(n)} \mid P \varepsilon \Omega\}$, of n-fold power probability distributions if

$$E_{P^{(n)}}\{h[T(X_1, \cdots, X_n)]\} = 0$$

Received October 7, 1959; revised April 18, 1960.

[1] This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-53). This paper in whole or in part may be reproduced for any purpose of the United States Government.

[2] The research was supported in part by a National Science Foundation Grant. A conversation with Professor Herman Rubin was helpful.

for all $P \varepsilon \Omega$ implies $h[T(x_1, \cdots, x_n)] = 0$, a.e., $P^{(n)}$, for all $F \epsilon \Omega$. The class Ω is said to be symmetrically complete whenever the latter condition holds.

Since the completeness of the order statistic plays an essential role in nonparametric estimation and hypothesis testing, e.g., Fraser [2] and Bell [1], it is of interest to determine those classes of *cpf's* for which the order statistic is complete.

Many of the traditionally studied classes of *cpf's* on the real line are known to be symmetrically complete, e.g., all continuous *cpf's* ([4], pp. 131–134, 152–153); all *cpf's* absolutely continuous with respect to Lebesgue measure ([3], pp. 23–31); and all exponentials of a certain form ([4], pp. 131–134).

The object of this note is to present a different ([4], pp. 131–134, 152–153) demonstration of the symmetric completeness of the class of all continuous *cpf's*; and to extend this and other known completeness results to probability spaces other than the real line, e.g., Fraser [2], and Lehmann and Scheffé [5], [6].

The paper is divided into four sections. Section 1 contains the introduction and summary. In Section 2 the notation and terminology are introduced. The main theorem is presented in Section 3, and some consequences of the proof of the main theorem and known results are indicated in Section 4.

2. Terminology and notation. Let $(X, \mathcal{S})$ be an arbitrary measurable space; λ, an arbitrary measure on $(X, \mathcal{S})$; and Ω, a class of probability measures on $(X, \mathcal{S})$.

Consistent with the notation of Scheffé [7] one defines the following sets and classes.

$\Omega_0(X)$ = the class of all probability measures on $(X, \mathcal{S})$;
$\Omega_1(X)$ = the class of all nondegenerate probability measures on $(X, \mathcal{S})$;
$\Omega_2(X)$ = the class of all nonatomic probability measures on $(X, \mathcal{S})$;
$\Omega_3(\lambda) = \{P \varepsilon \Omega_0(X) \mid P \ll \lambda\}$, i.e., the class of probability measures absolutely continuous with respect to λ;
$\Omega(\mathfrak{I}, \lambda) = \{\lambda_A \mid A \varepsilon \mathfrak{I}_\lambda^+\}$ where $\mathfrak{I}_\lambda^+ = \{A \varepsilon \mathfrak{I} \mid 0 < \lambda(A) < \infty\}$ and $\lambda_A(C) = \lambda(AC)/\lambda(A)$ for all $C \varepsilon \mathcal{S}$;
$\mathfrak{N}_\Omega = \{A \varepsilon \mathcal{S} \mid P(A) = 0 \text{ for all } P \varepsilon \Omega\}$, i.e., the null class of Ω;
$(X^{(n)}, \mathcal{S}^{(n)})$ = the product n-space generated by $(X, \mathcal{S})$;
$\lambda^{(n)} = \lambda \times \cdots \times \lambda$ = the n-fold power measure on $(X^{(n)}, \mathcal{S}^{(n)})$ generated by λ;
$\Omega^{(n)} = \{P^{(n)} \mid P \varepsilon \Omega\}$ = class of power measures generated by Ω;
$\mathfrak{N}_{\Omega^{(n)}} = \{A \varepsilon \mathcal{S}^{(n)} \mid P^{(n)}(A) = 0 \text{ for all } P \varepsilon \Omega\}$ = null class of $\Omega^{(n)}$.

A class Ω is said to be *symmetrically complete for* $n = k$ if $h_k = 0[P^{(k)}]$ i.e., $h_k = 0$ a.e. with respect to $P^{(k)}$, for all $P \varepsilon \Omega$, whenever h_k satisfies
(a) h_k is a symmetric function [measurable on $(X^{(k)}, \mathcal{S}^{(k)})$]; and
(b) $\int h_k \, dP^{(k)} = 0$ for all $P \varepsilon \Omega$.

With this notation we now demonstrate that the class $\Omega_2(X)$ is symmetrically complete for all n.

In the sequel it will be assumed that ν is an arbitrary fixed *nonatomic prob-*

ability measure on $(X, \mathcal{S})$; that h_n is a symmetric measurable function on $(X^{(n)}, \mathcal{S}^{(n)})$; and that $\mathcal{A}$ is a *semi-algebra* which generates $\mathcal{S}$. [Note: $\mathcal{A}$ is a semi-algebra if $X \varepsilon \mathcal{A}$; $\mathcal{A}$ is closed under finite intersections; and $A, B \varepsilon \mathcal{A}$ with $A \subset B$ implies the existence of $\{A_0, A_1, \cdots, A_m\} \subset \mathcal{A}$ such that $A = A_0 \subset A_1 \subset \cdots \subset A_m = B$ and $A_i - A_{i-1} \varepsilon \mathcal{A}$ for $i = 1, 2, \cdots, m$.]

3. The main theorem. The proof of the main theorem utilizes the facts that $\Omega(\mathcal{A}, \gamma)$ is symmetrically complete for properly chosen $\mathcal{A} \subset \mathcal{S}$; that the null classes of $\Omega^{(n)}(\mathcal{A}, P_1)$ and $\Omega_3^{(n)}(P_1)$ are equal; that, therefore, $\Omega_3(P_1)$ is symmetrically complete; and that so is $\Omega_2(X)$, since it is the union of classes $\Omega_3(P)$.

These ideas are given more precisely by the following three lemmas.

LEMMA 1. (Fraser) *If γ is an arbitrary nonatomic probability measure on $(X, \mathcal{S})$ and $\mathcal{A}$ is a semi-algebra which generates $\mathcal{S}$, then $\Omega(\mathcal{A}, \gamma)$ is symmetrically complete for all n.*

PROOF. See Fraser [2].

LEMMA 2. *If $P_1 \varepsilon \Omega_2(X)$, then $\mathfrak{N}_{\Omega^{(n)}(\mathcal{A}, P_1)} = \mathfrak{N}_{\Omega_3^{(n)}(P_1)}$ for all n.*

PROOF. Let n be an arbitrary fixed positive integer. Clearly, $P_1^{(n)}(A) = 0$ implies $P^{(n)}(A) = 0$ for all $P \varepsilon \Omega_3(P_1)$. This latter condition implies $\mathfrak{N}_{\{P_1^{(n)}\}} \subset \mathfrak{N}_{\Omega_3^{(n)}(P_1)}$. On the other hand, since

$$P_1^{(n)} \varepsilon \Omega^{(n)}(\mathcal{A}, P_1) \subset \Omega_3^{(n)}(P_1), \mathfrak{N}_{\{P_1^{n}\}} \supset \mathfrak{N}_{\Omega^{(n)}(\mathcal{A}, P_1)} \supset \mathfrak{N}_{\Omega_3^{(n)}(P_1)}.$$

The conclusion follows immediately.

The symmetric completeness of $\Omega(\mathcal{A}, P_1)$ and the equality of the two null classes are sufficient to establish the next lemma.

LEMMA 3. *If $P_1 \varepsilon \Omega_2(X)$, then $\Omega_3(P_1)$ is symmetrically complete for all n.*

PROOF. $\int h_n \, dP^{(n)} = 0$ for all $P \varepsilon \Omega_3(P_1)$ implies $P^{(n)}\{h_n \neq 0\} = 0$ for all $P \varepsilon \Omega(\mathcal{A}, P_1) \subset \Omega_3(P_1)$. Hence $\{h_n \neq 0\} \varepsilon \mathfrak{N}_{\Omega^{(n)}(\mathcal{A}, P_1)} = \mathfrak{N}_{\Omega_3^{(n)}(P_1)}$ and $h_n \equiv 0[P^{(n)}]$ for all $P \varepsilon \Omega_3(P_1)$.

The main theorem now follows from the preceding lemmas and the fact that any measure absolutely continuous with respect to a nonatomic measure is itself nonatomic.

MAIN THEOREM. *The class $\Omega_2(X)$ of all nonatomic probability measures on an arbitrary measurable space $(X, \mathcal{S})$ is a symmetrically complete class for all n. In particular, the class Ω_2 of all continuous cpf's on the real line is a symmetrically complete class for all n.*

PROOF. It is sufficient to demonstrate that for arbitrary fixed n, and arbitrary fixed $P_1 \varepsilon \Omega_2(X)$, $P_1^{(n)}\{h_n \neq 0\} = 0$, whenever h_n is a measurable symmetric function with the property: $\int h_n \, dP^{(n)} = 0$ for all $P \varepsilon \Omega_2(X)$.

Under such circumstances it is clear that $\Omega_3(P_1) \subset \Omega_2(X)$. Therefore, Lemma 3 guarantees for symmetric h_n such that $\int h_n \, dP^{(n)} = 0$ for all $P \varepsilon \Omega_2(X)$, that $P^{(n)}\{h_n \neq 0\} = 0$ for all $P \varepsilon \Omega_3(P_1)$. But $P_1 \varepsilon \Omega_3(P_1)$ and, consequently, $P_1^{(n)}\{h_n \neq 0\} = 0$.

4. Extensions. The symmetric completeness of several other classes of statistical interest can be extended to abstract spaces. In fact, by an extension of

the ideas above and those of Fraser ([2],[3], pp. 23–31), one can demonstrate the following result.

THEOREM. *If* $(X, \mathcal{S})$ *is an arbitrary measurable space, then* (I) $\Omega_0(X)$, $\Omega_1(X)$ *and* $\Omega_2(X)$ *are symmetrically complete for all n.*

If, further, λ *is a nonatomic,* σ*-finite measure on* $\mathcal{S}$ *and* $\mathcal{A}$ *is a semialgebra which generates* $\mathcal{S}$*, then,* (II) $\Omega(\mathcal{A}, \lambda)$, $\Omega(\mathcal{S}, \lambda)$ *and* $\Omega_3(\lambda)$ *are symmetrically complete for all n.*

REFERENCES

[1] C. B. BELL, "On the structure of distribution-free statistics," *Ann. Math. Stat.*, Vol. 31 (1960), pp. 703–709.

[2] D. A. S. FRASER, "Completeness of order statistics," *Can. J. Math.*, Vol. 6 (1954), pp. 42–45.

[3] D. A. S. FRASER, *Non-Parametric Methods in Statistics*, John Wiley and Sons, New York, 1957.

[4] E. L. LEHMANN, *Testing Statistical Hypotheses*, John Wiley and Sons, New York, 1959.

[5] E. L. LEHMANN AND HENRY SCHEFFÉ, "Completeness, similar regions and unbiased estimation," Part I, *Sankhyā*, Vol. 10 (1950), pp. 305–340.

[6] E. L. LEHMANN AND HENRY SCHEFFÉ, "Completeness, similar regions and unbiased estimation," Part II, *Sankhyā*, Vol. 15 (1955), pp. 219–236.

[7] H. SCHEFFÉ, "On a measure problem arising in the theory of non-parametric tests," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 227–233.

ON CENTERING INFINITELY DIVISIBLE PROCESSES[1]

BY RONALD PYKE

Stanford University and Columbia University

The concept of centering stochastic processes having independent increments, introduced by Lévy, is applied to processes having both stationary and independent increments. The main purpose of this note is to answer the question as to what centering functions preserve the stationarity of the increments.

In 1934, Lévy [1] proved that any stochastic process with independent increments may be transformed by subtraction of a sure function, called a centering function, into a process whose sample functions possess certain desirable smoothness properties. (cf. Lévy [2] and Doob [3]). It is clear that the transformed process, called the centered process, is also a process possessing independent increments. The purpose of this paper is to show that a process having stationary and independent increments may be centered in such a way so as to preserve the stationarity as well as the independence of the increments.

To be more precise, consider the following definitions (cf. Doob [3] p. 407).

Received September 28, 1959; revised December 15, 1959.

[1] This research was supported in part by the Office of Naval Research under contract Nonr-22521 (NR-042-993). Reproduction in whole or part is permitted for any purpose of the United States Government.

For a set $T \subset R_1$, let T^* denote the set of limit points of T except that the supremum and infimum of T are to be included in T^* only if they belong to T.

DEFINITION 1: A stochastic process $\{X_t : t \varepsilon T\}$ is said to be centered if and only if

(a) for every $\{t_n\} \subset T$ satisfying $t_n \nearrow t \varepsilon T^* (t_n \searrow t \varepsilon T^*)$ there exists a random variable $X_{t-}(X_{t+})$, independent of the particular sequence, such that

$$X_{t_n} \xrightarrow{\text{a.s.}} X_{t-}, \ (X_{t_n} \xrightarrow{\text{a.s.}} X_{t+})$$

(b) there exists a function g defined and continuous on the closure of T such that any difference $X_t - X_s$, $t, s \varepsilon T^*$, or any such difference with t replaced by $t+$ or $t-$ and/or s replaced by $s+$ or $s-$, is constant a.s. if and only if

$$X_t - X_s = g(t) - g(s) \text{a.s.}$$

(c) $X_{t-} = X_t = X_{t+}$ a.s. for all but at most a countable number of points of T.

This definition differs from that given by Doob only through condition (b). In Doob's definition, the function g was restricted to be constant over T^*. The above modified definition has the advantage of making it unnecessary to distinguish between degenerate and non-degenerate processes in the theorems below, as well as of insuring the truth of the statement that if $\{X_t : t \varepsilon T\}$ is a centered process, then so is $\{X_t + h(t) : t \varepsilon T\}$ for every continuous and bounded function h on T. This statement is not true under the more restrictive definition of Doob.

DEFINITION 2: A function $c: T \to R_1$ is said to be a centering function of a stochastic process $\{X_t : t \varepsilon T\}$ if and only if the process $\{X_t - c(t) : t \varepsilon T\}$ is centered.

It is clear that one may always find a centering function, such that the resulting centered process satisfies (b) of Definition 1 with $g = 0$.

A stochastic process, $\{X_t : t \varepsilon T\}$, having stationary and independent increments and for which $T = [0, +\infty)$ and $X_0 = 0$ a.s. is said to be an Infinitely Divisible (I.D.) process. As is evident from Lemma 1 below, a correspondence may be defined in a natural way between the class of infinitely divisible random variables (r.v.) and the class of I.D. processes. For properties of infinitely divisible r.v.'s used in this paper, the reader is referred to [4] and [5].

In the case of a centering function for processes with independent increments, uniqueness is clearly impossible. One possible centering function is that used by Doob ([3], p. 408), namely the solution to $E\{\arctan [X_t - c(t)]\} = 0$. It should be noted that this particular centering function would not preserve stationarity of increments in case the given process were a non-degenerate I.D. process.

Define for all $\omega \varepsilon R_1$ and $t \geqq 0$, $f(\omega : t) = E\{e^{i\omega X_t}\}$.

LEMMA 1: *A stochastic process* $\{X_t : t \geqq 0\}$ *having independent increments is an I.D. process if and only if there exist unique functions* $c: [0, \infty) \to R_1$ *and* $\psi : R_1 \to$ *complex plane, satisfying* (i) *for all* $s, t \geqq 0$, $c(s) + c(t) = c(s + t)$, (ii) *for all rational* $r \geqq 0$, $c(r) = 0$, (iii) *for all* $t \geqq 0$, $\omega \varepsilon R_1$, $\log f(\omega : t) = i\omega c(t) + t\psi(\omega)$.

PROOF: The proof of the sufficiency is left to the reader. The main interest is in the necessity of these conditions. A straightforward proof of this is possible using the Lévy-Khintchine representation of $f(\omega:t)$, namely

$$(1)\quad \log f(w:t) = i\omega\mu(t) + \int_{R_1} (e^{i\omega x} - 1 - i\omega x/(1 + x^2))(1 + x^2)/x^2\, dG(x:t).$$

where $G(\cdot:t)$ is a bounded non-decreasing right-continuous function, since clearly

$$(2)\qquad f(w:s + t) = f(w:s)f(w:t)$$

The purpose of the proof given here is to demonstrate that the powerful tool (1) is not essential for proving the necessity of the conditions of Lemma 1. This is important, it is felt, because the result stated as Lemma 1 should logically be proven very shortly after an I.D. process is defined, and because such a definition may well precede any discussion of infinitely divisible r.v.'s.

For each n, $[f(\omega:t)]^{1/n}$ is a characteristic function. It is well known, and easily proven, that therefore, for all $p > 0$, $[f(\omega:t)]^p$, properly defined, is a characteristic function and that for all $\omega \varepsilon R_1$ and $t \geqq 0$, $f(\omega:t) \neq 0$. Because of (2), $|f(\omega:s)|\,|f(\omega:t)| = |f(\omega:s + t)|$. Since $0 < |f(\omega, s)| \leqq 1$, the solution of this functional equation is given by $2 \log |f(\omega:t)| = t[\psi(\omega) + \psi(-\omega)]$ where $\psi(\omega) = \log f(\omega:1)$. Consequently, upon defining $q(\omega:t) = e^{t\psi(\omega)}[f(\omega:t)]^{-1}$, it follows that $|q(\omega:t)| = 1$, and that $q(\cdot:t)$ is a continuous function for each $t \geqq 0$. Moreover, since for rational r, $f(\omega:rt) = [f(\omega:t)]^r$, one has

$$q(\omega:t) = \lim e^{t\psi(\omega)}[f(\omega:t)]^{-rt} = \lim [f(\omega:1)/f(\omega:rt)]^t$$

where the limit is taken as $r \nearrow t^{-1}$ over the rationals. However, by (2),

$$f(\omega:1)/f(\omega:rt) = f(\omega:1 - rt)$$

is a characteristic function and hence so is $q(\omega:t)$. Because $|q(\omega:t)| = 1$, the proof is then complete since for each $t \geqq 0$, $q(\omega:t) = e^{i\omega c(t)}$ for some real number $c(t)$. It may be easily checked that the function c and ψ thus defined satisfy the required conditions.

By using the function c in Lemma 1, one obtains

COROLLARY 1: *For an I.D. process $\{X_t:t \geqq 0\}$, there exists a centering function c such that the resulting centered process $\{X_t - c(t):t \geqq 0\}$ is also an I.D. process.*

It is remarked that a stationarity preserving centering function for an I.D. process is unique up to the addition of straight lines through the origin.

It is evident that portions of Definition 1 are superfluous when applied to I.D. processes. In fact, one can easily prove

LEMMA 2: *An I.D. process is centered if and only if its characteristic function $f(\omega:t)$ is continuous in t.*

COROLLARY 2: *An I.D. process is centered if and only if for all sequences*

$$0 \leqq t_n \to t,\ X_{t_n} \xrightarrow{\text{a.s.}} X_t\,.$$

It is emphasized that the above results are neither difficult nor too surprising. The fact that a *centered* I.D. process has a characteristic function which satisfies $\log f(\omega:t) = t\psi(\omega)$ is well known (e.g., cf. Lévy [2] p. 186, Doob [3] p. 419, Ito [6]). The justification for the presentation of the above material is two-fold;

(i) Corollary 1 has not been located in the literature and

(ii) several recent papers in the literature indicate that Lemma 1 and Corollary 1 are not known.

Concerning (ii), several authors assume that (8) is true for all separable I.D. processes (cf. [7], [11]) while in other papers the exact role played by centering in the case of I.D. processes seems to have been misunderstood (cf. [8]). Furthermore, as a consequence of Lemma 1, the assumption (retaining the notation of the papers referred to) that $\phi(t:\lambda)$ be continuous in λ may be removed from Theorem 1 of [9] and from Theorem 1 of [10]. For example Theorem 1 (iii) of [9] could be strengthened to read: $F(x:\lambda) \in C_1$ if and only if $\phi(t:\lambda) = [f(t)]^{\lambda}e^{itc(\lambda)}$ where $f(t)$ is a characteristic function and where c is a function satisfying the conditions of Lemma 1.

As mentioned in the above paragraph separability is sometimes thought to imply that a process is centered. Although this is not true, it is possible to relate these two properties as well as the properties of measurability and of boundedness of sample functions, as stated in

LEMMA 3: *For a separable I.D. process* $X = \{X_t:t \geqq 0\}$, *the following conditions are equivalent*: (i) X *is centered*, (ii) X *is measurable*, (iii) *there exists a separating sequence which is a subset of the rational numbers*, (iv) *there exists an open interval in* $[0, \infty)$ *over which almost all sample functions are either bounded from above or below.*

REFERENCES

[1] P. LÉVY, "Sur les intégrales dont les éléments sont des variables aléatoires indépendotes," *Ann. Scuola Norm. d. Pisa*, (2) No. 3 (1934), pp. 337–366.

[2] P. LÉVY, *Théorie de l'Addition des Variables Aléatoires*, 2nd ed., Gauthier-Villars, Paris, 1954.

[3] J. L. DOOB, *Stochastic Processes*, J. Wiley and Sons, New York, 1953.

[4] B. V. GNEDENKO AND A. N. KOLMOGOROV, *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, Cambridge, Mass., 1954.

[5] M. LOÉVE, *Probability Theory*, D. Van Nostrand, New York, 1955.

[6] K. ITO, "On stochastic processes (I): infinitely divisible laws of probability," *Jap. J. Math.* Vol 18 (1942) pp. 261–301.

[7] G. BAXTER AND M. D. DONSKER, "On the distribution of the supremum functional for processes with stationary independent increments," *Trans. A.M.S.* Vol. 85, No. 1, (1957) pp. 73–87.

[8] V. E. BENÉS, "Characterization and decomposition of stochastic processes with stationary independent increments," (Abstract) Notices, *Ann. Math. Stat.*, Vol. 5 (1958), pp. 246–247.

[9] H. TEICHER, "On the convolution of distributions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 775–778.

[10] E. L. CROW, "A property of additively closed families of distributions," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 892–897.

[11] HERMAN RUBIN AND HOWARD G. TUCKER, "Estimating the parameters of a differential process," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 641–658.

THE STRONG LAW OF LARGE NUMBERS FOR A CLASS OF MARKOV CHAINS

By Leo Breiman[1]

University of California at Los Angeles

1. Introduction. The following problem has arisen in the study of Markov chains of the learning model type. (See [1] for definitions). Let the state space be, for example, the unit interval [0, 1] and let the chain have a unique invariant initial distribution $\pi(dx)$. Now let the chain be started at some point $x \in [0, 1]$; is it true that

$$(1) \qquad \frac{1}{N}\sum_{n=1}^{N} X_n \rightarrow E_\pi X_1 \quad \text{a.s.?}$$

From the ergodic theorem we know that there is a set $S \subset [0, 1]$ such that $\pi(S) = 1$, and, if $x \in S$, then (1) holds. In learning models, however, π may be singular with respect to Lesbesgue measure, so a stronger result is desirable. We prove for a wide class of chains, including learning models, that (1) holds for every possible starting point. This result is well known for chains satisfying Doeblin's condition. Unfortunately, learning models do not.

2. The theorem. Let the state space Ω be a compact Hausdorff space, and $\mathcal{B}$ the Baire σ-field in Ω. The Markov transition probabilities $P(A \mid x)$ are assumed probabilities on $\mathcal{B}$ for fixed x, $\mathcal{B}$-measurable functions on Ω for fixed A, and such that there is a unique probability π on $\mathcal{B}$ satisfying

$$\pi(A) = \int P(A \mid x)\pi(dx), \qquad \text{all } A \in \mathcal{B}.$$

Let C be the class of all continuous functions on Ω, and add the final restriction that, if $f \in C$, so is $E(f(X_1) \mid X_0 = x)$. Let $\Omega^{(\infty)}$ be the infinite sequence space with coordinates in Ω. In the usual way, we construct a σ-field $\mathcal{B}^{(\infty)}$ in $\Omega^{(\infty)}$ and, using the initial distribution $X_0 = x$, a probability P_x on $\mathcal{B}^{(\infty)}$. Then

THEOREM. *Let $\phi \in C$, Then, for any $x \in \Omega$,*

$$\frac{1}{N}\sum_{1}^{N} \phi(X_n) \rightarrow E_\pi \phi(X_1) \quad \text{a.s. } P_x.$$

PROOF. The proof of this theorem is a combination of the Kakutani-Yosida norms ergodic lemma and an argument concerning conditional probabilities.

3. The topological part. We prove first a proposition which summarizes the topological ergodic theorem we need. Define the operator T on C into C by $(T\phi)(x) = E(\phi(X_1) \mid X_0 = x)$, so that $(T^k\phi)(x) = E(\phi(X_k) \mid X_0 = x)$, and

Received October 16, 1959.

[1] This research was carried on during the tenure of a National Science Foundation Fellowship.

set $\bar{T}_N\phi = \sum_1^N T^n\phi/N$. Then

PROPOSITION 1. For any $\phi \varepsilon C$, $\bar{T}_N\phi$ converges uniformly to $E_\pi\phi(X_1)$.

PROOF. This proposition and its proof are well known in linear space theory. However, for completeness, we give a short demonstration. Let $\mathfrak{M}$ be the class of probability measures on $\mathfrak{B}$, and consider the operators V, $\bar{V}_N$ on $\mathfrak{M}$ into $\mathfrak{M}$ defined by

$$(VQ)(A) = \int P(A \mid x)Q(dx), \qquad \bar{V}_N Q = \sum_1^N V^k Q/N.$$

By the Helly-Bray theorem, $\mathfrak{M}$ is closed and compact in the weak dual topology, so that there are plenty of convergent subsequences $\bar{V}_{N_k}Q$. But every limit point of $\bar{V}_NQ$ is invariant under V and hence is identified with π, so that $\bar{V}_NQ \to \pi$ in our topology. Therefore, for every $Q \varepsilon \mathfrak{M}$, and $\phi \varepsilon C$ we have

$$(\bar{T}_N\phi, Q) = (\phi, \bar{V}_NQ) \to (\phi, \pi)$$

and hence $\bar{T}_N\phi$ converges weakly to $E_\pi\phi$. Applying the Kakutani-Yosida norms ergodic lemma (see, for example, [2], pg. 441), we conclude that $\bar{T}_N\phi$ converges uniformly to $E_\pi\phi$.

4. The probabilistic part. Let $X_1, X_2, \cdots$ be distributed according to P_x, and define

$$Z_n^{(1)} = \begin{cases} \phi(X_n) - E(\phi(X_n) \mid X_{n-1}), & n > 1 \\ 0, & n \leqq 1 \end{cases}$$

$$Z_n^{(k)} = \begin{cases} E(\phi(X_n) \mid X_{n-k+1}) - E(\phi(X_n) \mid X_{n-k}), & n > k \\ 0, & n \leqq k. \end{cases}$$

PROPOSITION 2. $N^{-1}\sum_{n=1}^N Z_{(n)}^{(k)} \to 0$ a.s. P_x.

PROOF. We use the following result ([2], pg. 387). Let $Y_1, Y_2, \cdots$ be a sequence of random variables such that $E(Y_n \mid Y_{n-1}, \cdots, Y_1) = 0$ and

$$EY_n^2 \leqq M < \infty,$$

all n. Then

$$\frac{1}{N}\sum_{n=1}^N Y_n \to 0 \quad \text{a.s.}$$

To apply this, note that

$$E(Z_n^{(k)} \mid Z_{n-1}^{(k)}, \cdots, Z_1^{(k)}) \\ = E(E(Z_n^{(k)} \mid X_{n-k}, X_{n-k-1}, \cdots, X_1) \mid Z_{n-1}^{(k)}, \cdots, Z_1^{(k)}),$$

and that, since the $X_1, X_2, \cdots$ form a Markov chain,

$$E(Z_n^{(k)} \mid X_{n-k}, \ldots) = E(Z_n^{(k)} \mid X_{n-k}) = 0.$$

Further, $E(Z_n^{(k)})^2 \leqq 2(\sup |\phi|)^2$, thus giving the proposition.

5. Conclusion of the proof. To complete the demonstration of the theorem, write

$$\phi(X_n) - E(\phi(X_n) \mid X_{n-k}) = Z_n^{(1)} + Z_n^{(2)} + \cdots + Z_n^{(k)}, \qquad n > k.$$

Thus, by proposition 2,

$$\lim_N \left| \frac{1}{N} \sum_{n=1}^{N} \phi(X_n) - \frac{1}{N} \sum_{n=k+1}^{N} E(\phi(X_n) \mid X_{n-k}) \right| = 0, \quad \text{a.s. } P_x.$$

Or, neglecting at most k terms,

$$\lim_N \left| \frac{1}{N} \sum_{n=1}^{N} \phi(X_n) - \frac{1}{N} \sum_{n=1}^{N} E(\phi(X_{n+k}) \mid X_n) \right| = 0, \quad \text{a.s. } P_x,$$

so that, for fixed M,

$$\lim_N \left| \frac{1}{N} \sum_{n=1}^{N} \phi(X_n) - \frac{1}{N} \sum_{n=1}^{N} \left[\frac{1}{M} \sum_{k=1}^{M} E(\phi(X_{n+k}) \mid X_n) \right] \right| = 0, \quad \text{a.s. } P_x.$$

By proposition 1, for any $\epsilon > 0$, we may choose M such that

$$\left| \frac{1}{M} \sum_{k=1}^{M} E(\phi(X_{n+k}) \mid X_n) - E_\pi \phi(X_1) \right| \leq \epsilon,$$

and for such an M we have

$$\lim_N \left| \frac{1}{N} \sum_{n=1}^{N} \phi(X_n) - E_\pi \phi(X_1) \right| \leq \epsilon \quad \text{a.s. } P_x)$$

proving the theorem.

REFERENCES

[1] Samuel Karlin, "Some random walks occurring in learning models," *Pacific J. Math.*, Vol. 3 (1933), 725–756.

[2] Michael Loève, *Probability Theory*, D. Van Nostrand, New York, 1955.

EMPTINESS IN THE FINITE DAM

By A. Ghosal

Central Fuel Research Institute, Dhanbad, India

1. Summary: The paper discusses the general problem of emptiness in the finite dam and considers the probability that, starting with an arbitrary storage, the dam dries up before it fills completely. Some exact results are given both for discrete and continuous inputs. An interesting relation between this probability and the asymptotic distribution function of the dam content has also been obtained.

Received April 9, 1959; revised February 20, 1960.

2. Introduction: This paper is based on the storage system model given by Moran [5]. The storage, Z_t, of a dam of finite capacity, k, is defined for discrete time t ($t = 0, 1, 2, \cdots$) as the dam content just after an instantaneous release at time t, and just before an input, X_t, flows into it over the time interval $(t, t + 1)$. The model is subject to the conditions

(i) the inputs X_t during the intervals $(t, t + 1)$ are independently and identically distributed;

(ii) there is an overflow, max $(Z_t + X_t - k, 0)$, during the interval $(t, t + 1)$, while min $(k, Z_t + X_t)$ is left in the dam just before the release occurs;

(iii) the amount of water released at time $t + 1$ is min $(m, Z_t + X_t)$, where m is a constant $< k$.

It has been shown that the processes (Z_t) and $(Z_t + X_t)$ are both Markov chains, and the problem of obtaining their stationary distributions has been dealt with by Moran [5], [6], Gani [2], Gani and Prabhu [3] and Prabhu [7], [8].

This paper deals with the problem of finding the probability that, given an arbitrary initial storage and the distribution of the input (X_t), the dam dries up before it fills completely. It also shows that this probability bears an elegant relationship with the asymptotic distribution function of the dam content. D. G. Kendall [4] derived the time required by an infinite dam to dry up; Prabhu [8] dealt with the probability of emptiness at a given storage level for the finite dam, but for $m = 1$. Here, the problem for any m is dealt with.

3. The Probability of Emptiness–Discrete Input: If the release rules given in Section 2 operate, we have

$$Z_{t+1} = \begin{cases} Z_t + X_t - m & \text{if } m < Z_t + X_t < k, \\ 0 & \text{if } Z_t + X_t \leqq m, \\ k - m & \text{if } Z_t + X_t \geqq k. \end{cases} \tag{1}$$

Let $\{g_j\}$ be the probability distribution of X_t, so that

$$\Pr\{X_t = j\} = g_j, \qquad (j = 0, 1, \cdots). \tag{2}$$

Let V_i be the conditional probability that, starting with storage i, the dam becomes empty before it fills completely. It is easy to derive the following:

$$V_i = \begin{cases} \sum_{j=0}^{m-i} g_j + \sum_{j=1}^{k-m-1} g_{j+m-1} V_j & (i \leqq m), \\ \sum_{j=i-m}^{k-m-1} g_{j+m-1} V_j & (m < i \leqq k - m - 1). \end{cases} \tag{3}$$

We note that the states 0 and $k - m$ are absorbing, so that $V_i = 1$ for $i \leqq 0$, $V_{k-m+r} = 0$ for $r \geqq 0$.

3.1. *Geometric Input*: Consider an input distribution of the geometric type,

$$g_j = ab^j \qquad (b = 1 - a, j = 0, 1, \cdots). \tag{4}$$

Applying the transformation $V_i = 1 - b^{-i}\phi_i$ to (3), after substituting (4)

in (3), we get

$$\phi_i = \begin{cases} \alpha & (i \leq m), \\ \alpha - \lambda \sum_{j=1}^{i-m-1} \phi_j & (i > m), \end{cases} \tag{5}$$

where

$$\lambda = ab^m, \tag{6}$$

$$\alpha = b^k + \lambda \sum_{j=1}^{k-m-1} \phi_j .$$

We solve for ϕ_i successively for the ranges $(m, 2m)$, $(2m, 3m)$, $\cdots$ in terms of the unknown constant α. For instance, for $m < i \leq 2m$, we get

$$\phi_i = \alpha[1 - \lambda(i - m - 1)].$$

Let $k = (N + 1)m + U$, where $0 \leq U < m$. We get the general expression

$$\phi_i = \alpha \sum_{q=0}^{n} (-\lambda)^q \binom{i - qm - 1}{q}, \quad (nm < i \leq (n+1)m, \; n = 0, 1, \cdots N + 1). \tag{7}$$

We solve for α from (6) and (7):

$$\alpha = b^k \Big/ \left[\sum_{q=0}^{N+1} (-\lambda)^q \binom{k - qm - 1}{q} \right]. \tag{8}$$

From (7) we have

$$V_i = 1 - ab^{-i} \sum_{q=0}^{n} (-\lambda)^q \binom{i - qm - 1}{q} \quad (nm < i \leq (n+1)m; \; n = 0, 1, \cdots N). \tag{9}$$

In many cases, it may be enough to know the bounds within which V_i should lie, and these bounds are given by Feller ([1], inequalities 8.11, 8.12 on p. 303). Prabhu [8] has obtained the bounds for $m = 1$. For general m, if we put $E(U_i) = E(X_i - m) = \rho - m$, where ρ is the mean input, we have

$$\begin{aligned} &(Z_0^{k-1} - Z_0^{m+i-1})/(Z_0^{k-1} - 1) \leq V_i \leq 1 && (\rho < m), \\ &(Z_0^{m+i-1} - Z_0^{k-1})/(1 - Z_0^{k-1}) \leq V_i \leq Z_0^i && (\rho > m), \\ &1 - (m + i - 1)/(k - 1) \leq V_i \leq 1 && (\rho = m), \end{aligned} \tag{10}$$

where Z_0 is the unique positive root (other than unity) of the equation $\sum_j Z^j \Pr(U_i = j) = 1$, i.e. $\sum_{j=0}^{\infty} Z^j g_j = Z^m$, and $Z_0 \gtrless 1$ according as $\rho \gtrless m$.

4. Continuous Input: It would be instructive to study the continuous analogue of (3). If $V(y)$ is the continuous analogue of V_i, the equations (3) become

$$V(y) = \begin{cases} G(m - y) + \int_0^{k-m} V(t)\, dG(t + m - y) & (0 < y \leq m), \\ \int_{y-m}^{k-m} V(t)\, dG(t + m - y) & (m \leq y < k - m), \end{cases} \tag{11}$$

where $G(x) = \Pr(X_i \leq x)$.

4.1. *Exponential Input*: Consider an exponential input of the type

$$dG(x) = \mu e^{-\mu x}\, dx, \qquad (0 < x < \infty;\ \mu > 0). \tag{12}$$

By applying the transformation $V(t) = 1 - e^{\mu t}\phi(t)$ and substituting (12) in (11), we get

$$\phi(y) = \begin{cases} \alpha & (y \leqq m), \\ \alpha - \lambda \int_0^{y-m} \phi(t)\, dt & (y > m), \end{cases} \tag{13}$$

where

$$\begin{aligned} \lambda &= \mu e^{-\mu m} \\ \alpha &= e^{-\mu k} + \lambda \int_0^{k-m} \phi(t)\, dt. \end{aligned} \tag{14}$$

Suppose $k = (N + 1)m + U$, where $0 \leqq U < m$. We can solve for $\phi(y)$ successively for the ranges $(m, 2m)$, $(2m, 3m)$, etc.

For $nm < y \leqq (n + 1)m$, we have

$$\phi(y) = \alpha \sum_{q=0}^{n} (-\lambda)^q \frac{(y - qm)^q}{q!} \qquad (n = 0, 1, \cdots N + 1). \tag{15}$$

α is determined as follows:

$$\begin{aligned} \alpha &= e^{-\mu k} + \lambda \int_0^{k-m} \phi(t)\, dt \\ &= e^{\mu k} + \lambda \left[\int_0^m \phi(t)\, dt + \cdots + \int_{Nm}^{k-m} \phi(t)\, dt \right], \end{aligned}$$

so that

$$\alpha = e^{-\mu k} \bigg/ \sum_{q=0}^{N+1} \frac{(-\lambda)^q (k - qm)^q}{q!}. \tag{16}$$

Thus, we have

$$V(y) = \begin{cases} 1 - e^{\mu y}\alpha & (y \leqq m), \\ 1 - e^{\mu y}\alpha \sum_{q=0}^{n} (-\lambda)^q \dfrac{(y - qm)^q}{q!} & (nm < y \leqq (n + 1)m;\ n = 0, 1, \cdots N). \end{cases} \tag{17}$$

We have the boundary conditions: $V(0) = 1$, $V(k - m + r) = 0$ for $r \geqq 0$. From (17) we find $V(+0) = 1 - \alpha$ and $V(k - m - 0) > 0$, indicating that there are points of discontinuities at $y = 0$ and $y = k - m$.

4.2. *Gamma Input*: Consider a gamma input

$$dG(x) = (\mu^p/(p - 1)!)e^{-\mu x}x^{p-1}\, dx \quad (0 < x < \infty;\ \mu > 0;\ p = 1, 2, \cdots). \tag{18}$$

Again, applying the transformation $V(t) = 1 - e^{\mu t}\phi(t)$ to (11), we get

$$\phi(y) = \begin{cases} \sum_{\gamma=0}^{p-1} \frac{\alpha_\gamma y^\gamma}{\gamma!} & (y \leq m), \\ \sum_{\gamma=0}^{p-1} \frac{\alpha_\gamma y^\gamma}{\gamma!} - \lambda \int_0^{y-m} \phi(t) \frac{(y-m-t)^{p-1}}{(p-1)!}\, dt & (y > m), \end{cases} \tag{19}$$

where

$$\lambda = (-1)^{p-1}\mu^p e^{-\mu m}, \tag{20}$$

$$\alpha_\gamma = e^{-\mu k}(-\mu)^\gamma \sum_{s=0}^{p-\gamma-1} \frac{(\mu k)^s}{s!} + \mu^p e^{-\mu m}(-1)^\gamma \int_0^{k-m} \phi(t) \frac{(t+m)^{p-\gamma-1}}{(p-\gamma-1)!}\, dt \tag{21}$$

$$(\gamma = 0, 1, \cdots p-1).$$

We get

$$\phi(y) = \sum_{\gamma=0}^{p-1} \alpha_\gamma \sum_{q=0}^{n} (-\lambda)^q \frac{(y-qm)^{qp+\gamma}}{(qp+\gamma)!} \quad \begin{array}{r}(nm < y \leq (n+1)m; \\ n = 0, 1, \cdots N+1),\end{array} \tag{22}$$

where $k = (N+1)m + U$, as in Section 4.1.

Prabhu [7] obtained (22) while deriving the distribution of dam storage. The α_γ's can be obtained by his method ([7], eqn. (13)).

Finally, we have

$$V(y) = \begin{cases} 1 - e^{\mu y} \sum_{\gamma=0}^{p-1} \frac{\alpha_\gamma y^\gamma}{\gamma!} & (y \leq m), \\ 1 - e^{\mu y} \sum_{\gamma=0}^{p-1} \alpha_\gamma \sum_{q=0}^{n} (-\lambda)^q \frac{(y-qm)^{qp+\gamma}}{(qp+\gamma)!} & \begin{array}{l}(nm < y \\ \leq (n+1)m, n \geq 0).\end{array} \end{cases} \tag{23}$$

$V(y)$ has two points of discontinuity at $y = 0$, $y = k - m$ since the boundary conditions are $V(0) = 1$, $V(k - m + r) = 0$ for $r \geq 0$.

5. Relationship with the Asymptotic Distribution of Dam Content: If $H(y)$ is the stationary c.d.f. of the dam content $Z_t + X_t$ we get the following integral equation ([7], eqn. (2)) for continuous input:

$$H(y) = \begin{cases} -\int_m^{m+y} H(t)\, dG(m+y-t) & (y < k-m), \\ G(y-k+m) - \int_m^k H(t)\, dG(m+y-t) & (y \geq k-m). \end{cases} \tag{24}$$

By applying the transformation $H(k - y) = 1 - e^{\mu y}\phi(y)$ to the above, for exponential and gamma inputs, (12) and (18), we obtain the same integral equations in $\phi(y)$, (13) and (19), as were obtained by applying the transformation $V(y) = 1 - e^{\mu y}\phi(y)$ to the integral equations for $V(y)$. We, therefore, obtain

$$V(y) = H(k - y). \tag{25}$$

We may verify that (25) holds good for discrete input also.

6. Acknowledgments: I am indebted to N. U. Prabhu, D. G. Kendall and the referee for helpful suggestions, and to Sultana Z. Ali for a useful comment.

REFERENCES

[1] W. Feller, *An Introduction to Probability Theory and its Applications*, John Wiley and Sons, New York, 1957, 2nd Ed.

[2] J. Gani, "Problems in the probability theory of storage systems," *J. Roy. Stat. Soc. Ser. B.* Vol. 14 (1957), pp. 181–207.

[3] J. Gani and N. U. Prabhu, "Stationary distributions of the negative exponential type for the infinite dam," *J. Roy. Stat. Soc., Ser. B.* Vol. 19 (1957), pp. 342–351.

[4] David G. Kendall, "Some problems in the theory of dams," *J. Roy. Stat. Soc., Ser. B.*, Vol. 19 (1957), pp. 207–212.

[5] P. A. P. Moran, "A probability theory of dams and storage systems," *Aust. J. Appl. Sci.*, Vol. 5, (1954), pp. 116–124.

[6] P. A. P. Moran, "A probability theory of dams and storage systems; modifications of the release rules," *Aust. J. Appl. Sci.*, Vol. 6 (1955), pp. 117–130.

[7] N. U. Prabhu, "On the integral equation for the finite dam," *Quart. J. of Math.*, Vol. 9 (1958), pp. 183–188.

[8] N. U. Prabhu, "Some exact results for the finite dam," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 1234–1243.

CORRECTION NOTES

CORRECTION TO "THE INDIVIDUAL ERGODIC THEOREM OF INFORMATION THEORY"

By Leo Breiman

University of California at Los Angeles

Mr. James Abbott has pointed out that the argument on Page 811 of the above-cited work, *Ann. Math. Stat.*, Vol. 28, No. 3 (1957), pp. 809–811, is incorrect. The results of the paper are valid, however, and Page 811 may be replaced by the following discussion.

Note that

$$E\left(\frac{p(x_{-k}, \cdots, x_{-1})}{p(x_{-k}, \cdots, x_0)} \,\middle|\, x_0, \cdots, x_{-k+1}\right) \leqq \frac{p(x_{-k+1}, \cdots, x_{-1})}{p(x_{-k+1}, \cdots, x_0)}$$

with probability one. By the concavity of log, it follows that the g_k sequence,

$$g_k = -\log_2\left(\frac{p(x_{-k}, \cdots, x_0)}{p(x_{-k}, \cdots, x_{-1})}\right)$$

satisfies

$$E(g_k \mid x_0, \cdots, x_{-k+1}) \leqq g_{k-1}.$$

Since $g_k \geqq 0$, and $Eg_0 < \infty$, the g_k sequence forms a non-negative lower semi-martingale and hence converges a.s. Actually, the convergence of the g_k sequence has been previously established by McMillan in [2].

Now consider $P(\sup_{k \leqq n} g_k > \lambda)$, and define the disjoint sets

$$E_j = \{g_j > \lambda, \ \sup_{k<j} g_k \leqq \lambda\},$$

whence $P(\sup_{k \leqq n} g_k > \lambda) = \sum_{j=1}^{n} P(E_j)$. Let Z_i be the cylinder sets $\{x_0 = a_i\}$ and $f_k^{(i)}$ the functions $-\log_2 P(x_0 = a_i \mid x_{-1}, \cdots, x_{-k})$. If $\sum_A f(x_0, x_{-1}, \cdots)$ indicates the sum of $f(x_0, x_{-1}, \cdots)$ over all sequences $(x_0, x_{-1}, \cdots) \varepsilon A$, then

$$P(E_j) = \sum_{E_j} p(x_{-j}, \cdots, x_0) = \sum_i \sum_{E_j \cap Z_i} \frac{p(x_{-j}, \cdots, x_0)}{p(x_{-j}, \cdots, x_{-1})} p(x_{-j}, \cdots, x_{-1}).$$

But on E_j we have the inequality

$$\frac{p(x_{-j}, \cdots, x_0)}{p(x_{-j}, \cdots, x_{-1})} = 2^{-g_j} \leqq 2^{-\lambda},$$

leading to

$$P(E_j) \leqq 2^{-\lambda} \sum_i \sum_{E_j \cap Z_i} p(x_{-j}, \cdots, x_{-1}) = 2^{-\lambda} \sum_i P(f_j^{(i)} > \lambda, \sup_{k<j} f_k^{(i)} \leqq \lambda).$$

Finally, then,

$$P\,(\sup_{k\leqq n} g_k > \lambda) \leqq 2^{-\lambda} \sum_i P\,(\sup_{k\leqq n} f_k^{(i)} > \lambda) \leqq s\cdot 2^{-\lambda},$$

where s is the number of values that the process ranges over. This last inequality gives $P(\sup_k g_k > \lambda) \leqq s\cdot 2^{-\lambda}$, which quickly leads to $E(\sup_k g_k) < \infty$.

CORRECTION TO

"BOUNDS ON NORMAL APPROXIMATIONS TO STUDENT'S AND THE CHI-SQUARE DISTRIBUTIONS"

By David L. Wallace

University of Chicago

The following correction should be made on p. 1127 of the above-titled article (*Ann. Math. Stat.*, Vol. 30 (1959), pp. 1121–1130): In the conclusion of Corollary 2 to Theorem 4.2, the exponent of n should be $-\frac{1}{2}$ and not $\frac{1}{2}$.

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Stanford Annual Meeting of the Institute, August 23-26, 1960. Additional abstracts will appear in the December 1960 issue.)

1. Estimating the Infinitesimal Generator of a Finite State Continuous Time Markov Process. Arthur Albert.

Let $\{Z(t), t > 0\}$ be a separable, continuous time Markov Process with stationary transition probabilities $P_{ij}(t)$, $i, j = 1, 2, \cdots, M$. Under suitable regularity conditions, the matrix of transition probabilities, $P(t)$, can be expressed in the form $P(t) = \exp tQ$, where Q is an $M \times M$ matrix and is called the "infintesimal generator" for the process.

In this paper, a density on the space of sample functions over $[0, t)$ is constructed. This density depends upon Q. If Q is unknown, the maximum likelihood estimate

$$\hat{Q}(k, t) = \|\hat{q}_{ij}(k, t)\|,$$

based upon k independent realizations of the process over $[0, t)$ can be derived. If each state has positive probability of being occupied during $[0, t)$ and if the number of independent observations, k, grows large (t held fixed), then $\hat{q}_{ij}$ is strongly consistent and the joint distribution of the set $\{(k)^{\frac{1}{2}}(\hat{q}_{ij} - q_{ij})\}_{i \neq j}$ (suitably normalized), is asymptotically normal with zero mean and covariance equal to the identity matrix. If k is held fixed (at one, say) and if t grows large, then $\hat{q}_{ij}$ is again strongly consistent and the joint distribution of the set $\{(t)^{\frac{1}{2}}(q_{ij} - q_{ij})\}_{i \neq j}$ (suitably normalized), is asymptotically normal with zero mean and covariance equal to the identity matrix, provided that the process $\{Z(t)\ t > 0\}$ is metrically transitive (but not necessarily stationary) and has no transient states.

The asymptotic variances of the $\hat{q}_{ij}$ are computed in both cases.

2. The Sequential Design of Experiments for Infinitely Many States of Nature. Arthur Albert. (By title)

In a recent paper (*Ann. Math. Stat.* Vol. 30 (1959), pp. 755-770) Chernoff discussed a problem which he called "The Sequential Design of Experiments" as it applied to the two action (hypothesis testing) case. In that paper, a procedure was exhibited for which the risk was approximately $-c \log c/I(\theta)$, when θ is the true state of nature, $I(\theta)$ is an appropriately defined information number and c, the cost per experimental trial, is small. It was also shown that in order for some other procedure to do significantly better for some value of the parameter, it must do worse by an order of magnitude (as $c \to 0$) at some other value of the parameter. These results were obtained under the assumption that the parameter space is finite. In the present paper, the assumption of finiteness is dispensed with. The procedures proposed here are closely akin to Chernoff's procedure, and analogous (though slightly weaker) optimality properties are derived.

3. Maximal Independent Stochastic Processes. C. B. Bell, University of California, Berkeley. (By title)

R. Pyke (1958) asked: What is the maximum cardinality, M_a, of a family of independent random variables defined on an abstract space Ω of cardinality a? (1) For $a < \aleph_0$, an elementary counting process yields $M_a = [\log_2 a]$. (2) For $a = \aleph_0$, a construction and a result of E. Marczewski (*Colloq. Math.*, 1955) yield $M_{\aleph_0} = \aleph_0$. (3) $M\mathfrak{c} = 2^{\mathfrak{c}}$ follows from a result

of Kakutani and Oxtoby (*Ann. Math.*, 1950) for the real line. (4) For $a \geq \mathfrak{C}^n$, one notes that a subset of Ω has the cardinality of a cartesian product of n real lines. Consequently, an elementary construction provides $M_a \geq n \cdot 2^{\mathfrak{C}}$. (5) Following the method mentioned in (4) above and using the Generalized Continuum Hypothesis it is established that $M_{\aleph_r} \geq \max [\aleph_2, \aleph_{r-1}]$ for ordinals r. *Open problem*: Can the Kakutani-Oxtoby construction be generalized to yield $M_a = 2^a$ for all $a > \mathfrak{C}$?

4. The Covariance Function of a Simple Trunk Group, with Applications to Traffic Measurement. V. E. Beneš, Bell Telephone Laboratories and Dartmouth College. (By title)

Erlang's classical model for telephone traffic is considered: N trunks, calls arriving in a Poisson process, and negative exponential holding-times. Let $N(t)$ be the number of trunks in use at t. An explicit formula for the covariance $R(\cdot)$ of $N(\cdot)$ in terms of the characteristic values of the transition matrix of the Markov process $N(\cdot)$ is obtained. Also, $R(\cdot)$ is expressed purely in terms of constants and the "recovery function," i.e., the transition probability $\Pr\{N(t) = N \mid N(0) = N\}$. $R(\cdot)$ is accurately approximated by $R(0) \exp \{r_1 t\}$, where r_1 is the largest negative characteristic value, itself well approximated (underestimated) by $-E\{N(\cdot)\}/R(0)$. Exact and approximate formulas for sampling error in traffic measurement are deduced from these results.

5. Limiting Distribution of the Maximum in an Infinite Sequence of Exchangeable Random Variables. Simeon M. Berman, Columbia University. (By title)

Let $\{X_n : n = 1, 2, \cdots\}$ be an infinite sequence of exchangeable random variables (r.v.'s), i.e., the joint distribution function (d.f.) of any m of these r.v.'s does not depend on their subscripts but only on their number m. The limiting d.f. of $Z_n = \max (X_1, X_2, \cdots, X_n)$ is characterized; a necessary and sufficient condition for the convergence of the d.f. of Z_n is given under an assumption on the d.f. of X_1. Let $\Phi(x)$ be one of the three limiting d.f.'s of maxima of independent r.v.'s with a common d.f.; let $A(y)$ be any d.f. such that $\lim_{y \to 0+} A(y) = 0$ and $\lim_{y \to \infty} A(y) = 1$. Then a d.f. is a limiting d.f. of Z_n if and only if is of the form $\int_0^\infty [\Phi(x)]^y \, dA(y)$. Let $\{Y_n : n = 1, 2, \cdots\}$ be independent r.v.'s with the same marginal d.f.'s as X_1; suppose that $W_n = \max (Y_1, Y_2, \cdots, Y_n)$ has the limiting d.f. $\Phi(x)$, that is, there exist sequences $\{a_n\}$ and $\{b_n\}$ such that for all t,

$$\lim_{n \to \infty} P\{a_n^{-1}(W_n - b_n) \leq t\} = \Phi(t).$$

Under this assumption a necessary and sufficient condition is given for

$$\lim_{n \to \infty} P\{a_n^{-1}(Z_n - b_n) \leq t\} = \int_0^\infty [\Phi(t)]^y \, dA(y).$$

For each integer k and real number u, let

$$\mu_k(u) = P\{X_1 > u, X_2 > u, \cdots, X_k > u\}[P\{X_1 > u\}]^{-k};$$

then $\{\mu_k(u) : k = 1, 2, \cdots\}$ is a moment sequence which uniquely determines a d.f. $A_u(y)$. The condition is that the d.f.'s $A_u(y)$ converge completely to $A(y)$ as $u \to \infty$.

6. Elements of the Sequential Design of Experiments. Stuart A. Bessler, Sylvania Electronic Defense Laboratories.

An experimenter observes a physical phenomenon, the outcome of which depends upon some unknown parameter θ belonging to a finite parameter space Θ. The experimenter

wishes to choose which of k-alternative hypothesis best describes the parameter θ. To aid him in his decision he may perform experiments, e, selected from an infinite experiment space $\mathfrak{F}$. At each stage of the experimental process the experimenter must either stop experimenting and choose a terminal action or continue experimenting in which case he must choose the next experiment. The "rule" which the experimenter uses in making these decisions will be called a sequential decision procedure. A sequential decision procedure is proposed and its optimal character is described. The procedure is demonstrated by applying it to the problem of choosing which of three normal populations with common variances has the largest mean. Several other examples are discussed. A measure of efficiency is defined, and for each example the efficiency of a common alternative decision procedure is computed.

7. Alias Sets of Error Vectors in the Theory of Error Correcting Group Codes.
R. C. BOSE, University of North Carolina and Case Institute of Technology.

Consider an $n \times r$ parity check matrix A, of rank r whose elements belong to the Galois field $GF(s)$, $s = p^m$. The letters of the code consist of all n-place row vectors γ for which $\gamma A = 0$. Suppose γ is transmitted over an s-ary channel, and the output is $\gamma + \epsilon$, $\epsilon = (e_1, e_2, \cdots, e_n)$. Then ϵ is the error vector. Let $\alpha_1, \alpha_2, \cdots, \alpha_n$ be the row vectors of A. The 2^r vectors for which $e_1\alpha_1 + e_2\alpha_2 + \cdots + e_n\alpha_n$ has a constant value may be said to form an alias set. Let Ω be the set of error vectors which we wish to correct with certainty. Then no alias set should contain more than one member from Ω. Subject to this condition one would like to maximize n for a given r. This principle is of very wide application. For example, let $s = 2$, and let Ω consist of all vectors with one non-zero or two adjacent non-zero coordinates. Then we get Abramson's (*IRE Trans.* Vol. IT5, 1959, pp. 150–157) single error and double adjacent error correcting (SEC-DAEC) code by choosing A such that the vectors $\alpha_1, \alpha_2, \cdots, \alpha_n, \alpha_1 + \alpha_2, \cdots, \alpha_{n-1} + \alpha_n$ constituting the set Ω^* are all distinct. For decoding we calculate $(\gamma + \epsilon)A = \epsilon A$. If ϵ belongs to Ω then ϵA belongs to Ω^*, and uniquely determines ϵ. The required condition is satisfied if $\alpha_i = (\beta_i, 1)$, $i = 1, 2, \cdots, 2^r - 1$, where β_i is the coefficient vector of the $(r-2)$th degree polynomial which represents the element x^i of $GF(2^{r-1})$, x being a primitive element.

8. On Methods of Constructing Sets of Mutually Orthogonal Latin Squares Using a Computer.
R. C. BOSE, I. M. CHAKRAVARTI, D. E. KNUTH, Case Institute of Technology and University of North Carolina. (Invited Paper)

This is in continuation of the work presented under the same title at the Midwestern Regional Meeting of IMS this year. The method is to start with module $G(2, 2t)$ whose elements are vectors $x = (a, b)$ where a is a residue class (mod 2) and b is a residue class (mod $2t$), the addition being defined by $(a_1, b_1) + (a_2, b_2) = (c, d)$ where $a_1 + a_2 = c$ (mod 2), $b_1 + b_2 = d$ (mod $2t$) and where $P_1[x_j] = a_j$ and $P_2[x_j] = b_j$ and $(0, 0), (0, 1), \cdots (0, 2t-1), (1, 0) \cdots (1, 2t-1)$ is the standard order. The existence of a set of m mutually orthogonal Latin squares based on a module G is known (Mann 1942) to be equivalent to the existence of a matrix $X_{m.4t} = ((x_{ij}))$ whose rows are elements of G and amongst the $4t$ differences of any two rows every element of G occurs once. The existence of $X_{m.4t}$ implies the existence of $A_{m.4t} = ((a_{ij})) = ((P_1[x_{ij}]))$ where $a_{ij} = 0$ or 1 and in every two-rowed submatrix of A the four possible pairs $\binom{0}{0}, \binom{0}{1}, \binom{1}{0}$ and $\binom{1}{1}$ occur as columns with equal frequency t. Starting with such a matrix $A_{m.4t}$ which exists whenever a Hadamard matrix of order $4t$ exists, a programme was written for adjoining a second coordinate b_{ij} to every a_{ij}, where b_{ij} belongs to the ring of residue classes (mod $2t$), so that a matrix $X_{m.4t} = (a_{ij}, b_{ij})$ could be obtained. For $t = 3$, this method yielded $m = 5$

mutually orthogonal Latin squares of order 12. These results have also been generalized in other directions for different orders.

9. **Best Fit to a Random Variable by a Random Variable Measurable with Respect to a σ-Lattice.** H. D. BRUNK, University of Missouri.

Let $(\Omega, \mathcal{S}, \mu)$ be a probability space and f a random variable. Let $\mathcal{L}$ be a sub-σ-lattice of $\mathcal{S}$. Then there is an $\mathcal{L}$-measurable random variable g (for real t, $\{\omega \varepsilon \Omega : g(\omega) < t\} \varepsilon \mathcal{L}$) minimizing $\int (f - g)^2 \, d\mu$ (if appropriate integrals exist) in the class of $\mathcal{L}$-measurable random variables. More generally, the squared difference may be replaced by the W. H. Young form $\Delta_\Phi(\cdot, \cdot)$ determined by an arbitrary convex function Φ: the $\mathcal{L}$-measurable random variable g minimizing $\int \Delta_\Phi(f, g) \, d\mu$ in the class of $\mathcal{L}$-measurable random variables is independent of Φ (assuming appropriate integrals exist). When $\mathcal{L}$ is a sub-σ-field of $\mathcal{S}$, then g is the conditional expectation $E(f \mid \mathcal{L})$. In special cases treated by van Eeden (*Indag. Math.*, Vol. 19(1957), pp. 128–136, 201–211) and by Ayer, Brunk, Ewing, Reid, Silverman, and Utz (*Ann. Math. Stat.*, Vol. 26(1955), pp. 641–647, 607–616, *Pac. J. Math.*, Vol. 7(1957), pp. 833–846) g is the solution of a problem in maximum likelihood estimation of ordered parameters; in these cases the σ-lattice $\mathcal{L}$ is not a σ-field.

10. **On the Non-null Distribution of the Studentized Difference between the Two Largest Sample Values** (Preliminary Report). ANDRÉ CROTEAU AND JACQUES ST-PIERRE, University of Montreal.

The non-null distribution of the difference between the two largest sample values has already been obtained by A. Zinger and J. St-Pierre (*Biometrika*, Vol. 45, Parts 3 and 4, December 1958, pp. 436–447) in the case of normal populations with known variances. In the case of unknown variances, the distribution of the studentized difference between the two largest sample values is obtained for three normal populations. The distribution takes the form of an iterated integral involving recurrence relations leading rather easily to numerical evaluations. A generalisation in the case of "n" populations is presently studied by the authors.

11. **Random Noise in Relay Control Systems.** R. C. DAVIS, Convair Division of General Dynamics Corporation.

A general method is developed to obtain the probability distribution of the error in a single closed-loop relay control system in which one controls a linear time-invariant dynamic element in the presence of a time-varying signal perturbed additively by Gaussian noise. The noise is allowed to be of a particular nonstationary type and specifically is the output of a perfect amplifier with time variable gain in cascade with a linear time-invariant filter with a rational amplitude versus frequency response—the input being the derivative of a Wiener process. The method used is the development of the theory of a particular type of discontinuous Markoff process for which the corresponding analogy in heat conduction is the conduction of heat in a moving medium in which there is a surface of discontinuity in medium velocity. In this way both the transient and steady state probability distributions of error are obtained. The probability distribution of error is obtained explicitly and involves line integrals of the Gaussian probability density function in the phase space of the error and certain of its time derivatives.

12. **Sample Size for a Specified Width Confidence Interval on the Variance of a Normal Distribution.** FRANKLIN A. GRAYBILL AND ROBERT D. MORRISON, Oklahoma State University. (By title)

If an experimenter decides to use a confidence interval to locate a parameter, he is concerned with at least two things: (1) Does the interval contain the parameter? (2) How wide

is the interval? In general, the answer to these questions cannot be given with absolute certainty, but must be given with a probability statement. The problem the experimenter then faces is the determination of the sample size n such that (A) the probability will be equal to $1 - \alpha$ that the confidence interval contains the parameter, and (B) the probability will be equal to β^2 that the width of the confidence interval will be less than d units (where α, β^2, and d are specified). $1 - \alpha$ will be called the confidence coefficient, and β^2 will be called the width coefficient. To solve this problem will generally require two things: (1) The form of the frequency function; (2) Some previous information on the unknown parameters. This suggests that the sample be taken in two steps; the first sample will be used to determine the number of observations to be taken in the second sample so that (A) and (B) will be satisfied. For a confidence interval on the mean of a normal population with unknown variance this problem has been solved by Stein for $\beta^2 = 1$. The purpose of this paper is to illustrate a method for determining n to satisfy (A) and (B) for the variance of a normal distribution. A set of tables is presented to which will be needed for the solution of this problem.

13. On the Unbiasedness of Yates' Method of Estimation Using Interblock Information. Franklin A. Graybill and V. Seshadri, Oklahoma State University. (By title)

In a balanced incomplete block model with blocks and errors random normal variables, Yates has shown that there are two independent unbiased estimates for any treatment contrast. These are referred to as intrablock and interblock estimators. Yates has also given a method for combining these two estimators which depend on the variance (unknown) and has shown how to estimate the variances from an analysis of variance. Since this combined estimator is used quite extensively, it seems desirable to study its properties. Graybill and Weeks have shown that Yates' combined estimator is based on a set of minimal sufficient statistics and have presented an estimator which is unbiased. *The purpose of this note is to show that Yates' estimator, which is based on intrablock and interblock information, is unbiased.*

14. On the Distribution of the Ratio of the Largest of Several Chi-Squares to an Independent Chi-Square with Application to Ranking Problems. S. S. Gupta and M. Sobel, Bell Telephone Laboratories.

The distribution of $\chi^2_{\max}/\chi^2_0$ and its upper percentage points are considered where $\chi^2_{\max}$ is the maximum of p independent chi-squares and χ^2_0 is a chi-square independent of the p others. A common number ν of degrees of freedom is the principal case considered and tables of percentage points (25%, 10%, 5%, 1%) are given for $\nu = 2(2)50$ and $p = 1(1)10$; the case $p = 1$ which reduces to an F-distribution being used as a check. The computed tables have an application in the selection of a subset containing the "best" of several Gamma or Type III populations, i.e., the one with the largest scale parameter. In particular, if several exponential populations are individually observed until exactly r failures are obtained from each then the above tables can be used for selecting a subset containing the one with the largest mean life.

15. Expected Values of Normal Order Statistics. H. Leon Harter, Wright-Patterson Air Force Base.

A brief history is given of the development of the theory of order statistics and of past efforts to tabulate their expected values for samples from a normal population. A fuller account is given of the method of computation of a five-decimal-place table of the expected values of all order statistics for samples of size n from a normal population. Included is

such a table for $n = 2(1)100$ and for values of n, none of whose prime factors exceeds seven, up through $n = 400$. Also included is a discussion of an approximation proposed by Blom, and a table of values of the constant α required for this approximation for selected values of n, together with interpolation formulas for estimating α for other values of n. A discussion is given of actual and potential uses of the tables.

16. **Circular Error Probabilities.** H. LEON HARTER, Wright-Patterson Air Force Base. (By title)

A problem which often arises in connection with the determination of probabilities of various miss distances of bombs and missiles is the following: Let x and y be two normally and independently distributed orthogonal components of the miss distance, each with mean zero and with standard deviations $\sigma_x \geqq \sigma_y$. Now for various values of $c = \sigma_y/\sigma_x$, it is required to determine (1) the probability P that the point of impact lies inside a circle with center at the target and radius $K\sigma_x$, and (2) the value of K such that the probability is P that the point of impact lies inside such a circle. Solutions of (1), for $c = 0.0(0.1)1.0$ and $K = 0.1(0.1)5.8$, and (2), for the same values of c and $P = 0.5, 0.75, 0.9, 0.95, 0.975, 0.99, 0.995, 0.9975$, and 0.999, are given, along with some hypothetical examples of the application of the tables.

17. **Comparison of Normal Scores and Wilcoxon Tests.** J. L. HODGES, JR. AND E. L. LEHMANN, University of California, Berkeley. (By title)

The normal scores test (i.e. the Fisher-Yates-c_1-test or the van der Waerden X-test) and the Wilcoxon test have been proposed for testing the equality of two distributions against the "shift" alternative that the populations have distributions $F(x)$ and $F(x - \theta)$. From the known limiting behavior of the test statistics one obtains an expression for the asymptotic relative efficiency $e(F)$ of Wilcoxon to normal scores. It is shown that $0 \leqq e(F) \leqq 6/\pi$ for all F, and that all values including the endpoints may be attained.

18. **Minimal Sufficient Statistics for the Two-Way Classification Mixed Model Design.** ROBERT A. HULTQUIST AND FRANKLIN A. GRAYBILL, Oklahoma State University. (By title)

A theorem proved by Rao and Blackwell reveals the importance of minimal sufficient statistics in point estimation problems. This theorem states: If Y is a vector of observations, S is a minimal sufficient statistic for a vector of parameters θ and $f(Y)$ is an unbiased estimate of $g(\theta)$, then $f(S) = E[f(Y) \mid S]$ is also an unbiased estimate of $g(\theta)$ based on S and such that variance $f(Y) >$ variance $h(S)$. We thus see that if we are interested in determining minimum variance unbiased estimators of variance components these estimators must be based on a minimal sufficient statistic. The objective of this paper is to exhibit minimal sufficient statistics for the *two-way classification mixed model* design with *unequal* numbers in the subcells.

19. **Three-Quarter Replicates of 2^3 and 2^4 Designs.** PETER W. M. JOHN, California Research Corporation.

Half replicates of 2^3 and 2^4 designs do not enable all the main effects and two-factor interactions to be estimated clear of two-factor interactions. Three-quarter replicates are obtained which give all main effects and two-factor interactions clear for the 2^4 design; for the 2^3 design main effects are clear and, if any one of the two-factor interactions is negligible,

the other two are clear. In each case, the effects are estimated by extracting half replicates from the design.

20. On the Generalization of Sverdrup's Lemma and Its Applications to Multivariate Distribution Theory. D. G. KABE, Karnatak University. (By title). (Introduced by B. D. Tikkiwal)

Tikkiwal and Kabe (*Karnatak Univ. J.*, 1958) have given analytic-cum-geometric proof of the Sverdrup's lemma (*Skand. Aktur.* 1947). This lemma is now generalized for a p-variate population. Let the vectors $X'_i = (x_{i1}, x_{i2}, \cdots, x_{in})$ for $i = 1, 2, \cdots, p$ have the density $f(X'_1X_1, X'_1X_2, \cdots, X'_pX_p, BX_1, BX_2, \cdots, BX_p)$, then the density of $X'_iX_j = b_{ij}$, $BX_i = v_i (i, j = 1, 2, \cdots, p)$, B being $q \times n$ matrix of rank q, is given by

$$2^{-p} \prod_{i=1}^{p} C(n - q - p + i) | BB' |^{-\frac{1}{2}p} f(b_{11}, b_{12}, \cdots,$$
$$b_{pp}, v_1, \cdots, v_p) | b_{ij} - v'_i(BB')^{-1}v_j |^{-\frac{1}{2}(n-q-p-1)}$$

$C(n)$ being the surface area of a unit n-dimensional sphere. Almost all the distributions in multivariate theory have been derived by the help of this lemma.

21. Approximations to Neyman Type A and Negative Binomial Distributions in Practical Problems (Preliminary Report). S. K. KATTI, Florida State University.

The Neyman Type A and the Negative Binomial distributions have been used for fitting data arising from biological phenomena with varying degrees of success, e.g. G. Beall, "The Fit and Significance of Contagious Distributions when Applied to Observations on Larval Insects", *Ecology*, Vol. 21 (1940), pp. 460–474 and C. I. Bliss and R. A. Fisher, "Fitting of Negative Binomial Distribution to Biological Data", *Biometrics*, Vol. 9, pp. 176–200. In the present work, it is shown that these distributions approximate to elementary distributions such as Poisson, Poisson with zeros added and Logarithmic in various regions of the parameter space. Preliminary fitting indicates that the elementary—and hence simple—distributions can be used with advantage as alternatives to these relatively complex distributions in many a practical situation. It is found that a reasonable judgment about the elementary distribution to be used can be made on the basis of the mean and the first frequency.

22. Two Sample Nonparametric Tests for Scale Parameter (Preliminary Report). JEROME KLOTZ, University of California, Berkeley.

Let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ be samples from populations with continuous distritions F and G. We are interested in tests of the hypothesis $F = G$ that will be powerful against differences in scale when the populations are equivalent in location. Siegel and Tukey have recently devised a way to use the Wilcoxon statistic for this problem. The Pitman efficiency of their test for the normal case relative to the F-test is $6/\pi^2$. Pitman efficiency one relative to the F can be obtained for normality with the use of the following rank order statistic S. Assign weight $[\Phi^{-1}(i/N + 1)]^2$ to the ith smallest observation in the pooled sample where $N = m + n$, and let S be the sum of these weights over the observations from F. (Our weights are the squares of those used in Van der Waerden's X-test for the corresponding location problem.) As a rank order statistic S has an exact null distribution; we give a small table for the null distribution and approximations for m, n large. The Pitman efficiency of the test of Siegel and Tukey relative to the S-test can take on any value between 0 and ∞ for different F.

23. Zero Correlation and Independence. H. O. LANCASTER, University of Sydney. (By title)

Let $\{x_j\}$ be a set of n random variables. Let orthonormal functions be defined on each marginal distribution such that $x_j^{(0)} = 1$ for $j = 1, 2, \cdots, n$ and so that $\{x_j^{(i_j)}\}$ forms a basis if x_j has only a finite number (n_j) of points of increase, $i_j = 1, 2, \cdots, n_j$ and $\{x_j^{(i_j)}\}$ is a complete orthonormal set if x_j has a general type of distribution. Let generalised coefficients of correlation be defined, $\rho^{\{i_j\}} = E\{\prod_j x_j^{(i_j)}\}$. These coefficients are ordinary coefficients of correlation if precisely two of the i_j are non-zero. If more than two of the i_j are non-zero, the generalized coefficients may not be less than unity in absolute value. *Theorem*: A necessary and sufficient condition for independence of the set $\{x_j\}$ is that the generalized coefficients of correlation should all vanish. The bivariate case has been treated by Sarmanov, O. V., *Doklady Akad. Nauk SSSR*, Vol. 121 (1958), pp. 52–55 and Lancaster, H. O., *Aust. J. Stat.*, Vol. 1 (1959), pp. 53–56 and *J. Aust. Math. Soc.* (in the press), in which the multivariate case of the theorem also is proved without restriction.

24. Sequential Model Building for Prediction in Regression Analysis, I (Preliminary Report). HAROLD J. LARSON AND T. A. BANCROFT, Iowa State University.

Two different sequential procedures for deciding on the "length" of the linear regression model to use for predictions are evaluated, both assuming the population variance σ^2 to be known. In the first procedure the experimenter fits all the independent variables available, sequentially tests the coefficients of the "doubtful" ones to be zero, and deletes from the model the terms whose coefficients do not differ significantly from zero. In the second procedure the experimenter fits a set of "basic" variables he knows to be necessary, sequentially test the coefficients of the "nonbasic" variables to be zero and adds to his model those nonbasic variables whose coefficients differ significantly from zero. The expected value and the variance of the estimator are discussed for each procedure and limited tables for certain specific values of the parameters are presented to allow explicit evaluation of the bias of the estimators.

25. The Use of Sample Quasi-Ranges in Setting Confidence Intervals for the Population Standard Deviation. F. C. LEONE, Y. H. RUTENBERG, AND C. W. TOPP,* Case Institute of Technology and* Fenn College.

The problem is the choice of an optimal selection method of quasi-ranges for setting one sided confidence bounds and confidence intervals for the standard deviation from a given distribution. The proposed methods of optimal selection are applied to random ordered samples from the normal, exponential and rectangular distributions. Tables of confidence bounds for the standard deviation of these distributions are given for confidence levels commonly used in statistical work. These are compared with the results of standard procedures.

26. On a Property of a Test for the Equality of Two Normal Dispersion Matrices Against One-sided Alternatives. WADIE F. MIKHAIL, University of North Carolina.

The monotonic character, with respect to the variation of each noncentrality parameter, of the power function of the largest root test of normal multivariate analysis of variance or of independence between two sets of variates was proved in an earlier paper by S. N. Roy

and the author. This paper obtains, using the same technique used before, similar results for four tests derived by S. N. Roy and R. Gnanadesikan for the equality of two dispersion matrices, in the normal multivariate set-up, against one-sided alternatives.

27. A Note on Simple Sampling Plans (Preliminary Report). T. V. NARAYANA AND S. G. MOHANTY, Queen's University.

From previous work done by one of the authors, it is known that a simple sampling plan of size n can be characterised by a unique vector of n non-negative integers satisfying certain conditions. A simple symmetric sampling plan of size n is defined as one in which the boundary points are symmetric about the line $y = x$. The following theorem is proved: The number of simple symmetric sampling plans of size n is $\binom{[3n/2]}{[(n-1)/2]} \Big/ [(n+1)/2]$ where $[x]$ is the largest integer contained in x. This theorem follows from known results on the number of compositions of an integer dominated by a given composition of this integer (*Canad. Math. Bull.*, Vol. 1, No. 3). A recursive method is suggested to obtain the number of simple sampling plans of size n and the authors hope to establish the general result that the number of such sampling plans is $\frac{1}{n}\binom{3n}{n-1}$.

28. On Sampling with Varying Probabilities and With Replacement in Sub-sampling Designs. J. N. K. RAO, Iowa State University. (Introduced by T. A. Bancroft)

In sub-sampling, it is usual practice to select the primaries with replacement and with varying probabilities, due to difficulties in the theory of sampling with varying probabilities and without replacement. This leads to three different methods of selecting the secondaries. In method 1, if the ith primary is selected λ_i times, $m_i\lambda_i$ secondaries are selected without replacement and with equal probabilities from the ith primary. In method 2, if the ith primary is selected λ_i times, λ_i sub-samples of size m_i are independently drawn of each other from the ith primary with equal probability and without replacement, each sub-sample being replaced after it is drawn. In method 3, when the ith primary is selected λ_i times, a fixed size of m_i is drawn from the ith primary with equal probability and without replacement and the estimate from the ith primary is weighted by λ_i. It is known that method 1 has smaller variance than method 2, and method 2 has smaller variance than method 3. But, the three methods have different expected costs, assuming that expected cost in a primary as proportional to expected sample size from the primary. Therefore it would appear more reasonable to compare the efficiency of the three methods for the same expected sample size. Here a comparison of the variances has been made for the same expected sample size but the conclusions remain the same regarding efficiency.

29. Some Results on Transformations in the Analysis of Variance. M. M. RAO, Carnegie Institute of Technology.

The square-root and the logarithmic transformations are considered when the mean is large in each case. In the former the variance is assumed known, and in the latter the corresponding assumption is that the coefficient of variation is small but the variance is unknown. In these cases, it is shown that the usual normal theory is applicable to test the hypotheses on means of the untransformed variables. These results extend those of E. G. Olds and N. C. Severo (These *Annals*, 1956). Sufficient conditions for the applicability of the normal theory are presented for a class of distributions depending on a finite set of

parameters with one parameter large, while the others, if any, are relatively small, or are confined to a fixed bounded set in the parameter space.

30. Normal Approximation to the Chi-square and Non-central F Probability Functions. NORMAN C. SEVERO AND MARVIN ZELEN, University of Buffalo and National Bureau of Standards. (By title)

Let x_p denote the $100p\%$ percentage point of the normal distribution, i.e., $\Phi(x_p) = \int_{-\infty}^{x_p} (2\pi)^{-\frac{1}{2}} \exp(-\frac{1}{2}t^2)dt = 1 - p$. It is shown that the $100p\%$ percentage point of the Chi-square distribution with ν degrees of freedom may be approximated by

$$\chi_p^2(\nu) \doteq \nu\{1 - [2/(9\nu)] + (x_p - h_\nu)[2/(9\nu)]^{\frac{1}{2}}\}^3$$

where h_ν is an auxiliary function whose value may be obtained by linear interpolation in one of two short tables (one for $\nu \geqq 30$, and one for $5 \leqq \nu < 30$) consisting of only 15 entries each. For values of p between .005 and .995, this improved "Wilson-Hilferty" approximation gives results in error by at most .01 for $\nu \geqq 30$, and at most .05 for $5 \leqq \nu < 30$. Let $P(F' \mid \nu_1, \nu_2, \lambda)$ denote the probability distribution function of the non-central F distribution with degrees of freedom ν_1 and ν_2, and non-centrality parameter λ. It is shown that $P(F' \mid \nu_1, \nu_2, \lambda) \doteq \Phi(x)$, where

$$x = \frac{\left\{\left(\frac{\nu_1 F'}{\nu_1 + \lambda}\right)^{1/3}\left(1 - \frac{2}{9\nu_2}\right) - \left(1 - \frac{2(\nu_1 + 2\lambda)}{9(\nu_1 + \lambda)^2}\right)\right\}}{\left\{\left[\frac{2(\nu_1 + 2\lambda)}{9(\nu_1 + \lambda)^2} + \frac{2}{9\nu_2}\left(\frac{\nu_1 F'}{\nu_1 + \lambda}\right)^{2/3}\right]^{1/2}\right\}}$$

For values of the parameters investigated, the error of the approximation is at most .01.

31. On a Geometrical Method of Construction of Cyclic PBIB (Preliminary Report). ESTHER SEIDEN, Northwestern University. (By title)

An effective method of construction of cyclic PBIB is found provided that the number of treatments is $2^{2m} - 1$, m a positive integer. Using the notation of R. C. Base and T. Shimamoto ("Classification and analysis of partially balanced incomplete block designs with two associate classes," *J.A.S.A.*, Vol. 47, 1952, the parameters are as follows: $v = 2^{2m} - 1$ $b = (2^m + 2)(2^m + 1)/2$ $r = (2^m + 2)/2$ $k = 2^m - 1$ $\lambda_1 = 1$ $\lambda_2 = 0$ $n_1 = 2^{2m}/2 - 2$ $n_2 = 2^{2m}/2$ $\alpha = 2^{2m}/4 - 3$ $\beta = 2^{2m}/4 - 1$ $v = 2^{2m} - 1$ $b = 2^m(2^m - 1)/2$ $r = 2^m/2$ $k = 2^m + 1$ $\lambda_1 = 1$ $\lambda_2 = 0$ $n_1 = 2^{2m}/2$ $n_2 = 2^{2m}/2 - 2$ $\alpha = 2^{2m}/4$ $\beta = 2^{2m}/4$. The construction makes use of the fact that in a projective plane with $2^m + 1$ points on a line there exists an effective construction of a Desarguessian plane based on a set of $2^m + 2$ points of which no three are on one line. The problem whether such a construction is possible if based on a non-Desarguessian plane is under investigation.

32. Distribution of Quantiles in Samples from a Bivariate Population. M. M. SIDDIQUI, Boulder Laboratories.

Let $F(x, y)$ be the distribution function of (X, Y) possessing a pdf $f(x, y)$. Let $F_1(x)$ (pdf $f_1(x)$) and $F_2(y)$ (pdf $f_2(y)$) be the marginal distributions of X and Y respectively. Given two numbers F_1 and F_2 in $(0, 1)$ let α and β be the numbers such that $F_1(\alpha) = F_1$ and $F_2(\beta) = F_2$. Assume that the first and second partial derivatives of $F(x, y)$ are continuous at (α, β) and $f(\alpha, \beta) \neq 0$.

A random sample (X_k, Y_k), $k = 1, \cdots, n$ is drawn and the values of X and Y are ordered separately so that $X'_1 < X'_2 \cdots < X'_n$; $Y'_1 < Y'_2 \cdots < Y'_n$. Let i and j be the integers such that $i/n \leq F_1 < (i + 1)/n$, $j/n \leq F_2 < (j + 1)/n$. Let M be the number of sample points (X, Y) such that $X < X'_i$ and $Y < Y'_j$. The joint distribution of (M, X'_i, Y'_j) is obtained and it is shown that it is asymptotically trivariate normal. The asymptotic correlation coefficient between (X'_i, Y'_j) is given by

$$\rho = \{(F - F_1F_2)/[F_1F_2(1 - F_1)(1 - F_2)]^{\frac{1}{2}}\}, \qquad F = F(\alpha, \beta).$$

The statistic M/n has asymptotic mean F and variance of order n^{-1}. This is used to set up confidence limits on ρ. A generalization to the asymptotic distribution of a set of quantiles in samples from a multivariate population is stated.

33. Power Characteristics of the Control Chart for Means. Frederick A. Sorensen, United States Steel Corporation Applied Research Laboratory.

Methods are derived for the determination of the Type I error probability and the power of the control chart for sample means (no standard given). Under the null hypothesis, the process is assumed to be $N(\mu, \sigma^2)$, where μ and σ are unknown constants. Under the alternatives considered, the process is assumed to be $N(\mu_i, \sigma^2)$ during the time interval from which the ith subgroup ($i = 1, \cdots, k$) is taken. For $k = 2, 3, 5, 10$ and 25, and subgroup sizes of 5 and 10, the power is tabulated with respect to two particular types of alternative believed to be typical of those encountered in practice: (1) One of the μ_i differs from the rest by an amount $\delta\sigma$ (single slippage); (2) Two of the μ_i differ from the rest by an equal amount $\delta\sigma$, but in opposite directions (symmetrical double slippage). The effect of using variable-width limits that produce a constant Type I error probability of 0.05 rather than using the traditional "three-sigma" limits is investigated. The power of the control chart is compared with that of the corresponding Model I analysis of variance test.

34. A Set of Sufficient Statistics for Variance Components in a Two-Way Classification Model With Unequal Numbers in the Subclasses. David L. Weeks and Franklin A. Graybill, Oklahoma State University. (By title)

One of the important methods of estimating variance components is by the analysis of variance (A.O.V.). The analysis of variance (A.O.V.) method of estimating variance components consists of obtaining an analysis of variance table, equating observed mean squares to expected mean squares and solving these equations for the estimates of the variance components. If the model is Eisenhart's Model II, then the A.O.V. method of estimating variance components gives estimators which are unbiased. If the model also has equal numbers in all subclasses, and all random variables are normally and independently distributed, the A.O.V. method gives unique, minimum variance, unbiased estimators. If the model is Eisenhart's Model II with equal numbers in all subclasses, and if all random variables are independently but not necessarily normally distributed, then the A.O.V. method of estimation gives estimators which are minimum variance, quadratic unbiased. However, if the model has unequal numbers in the subclasses, the problem is more complex. The A.O.V. method of estimation does not give minimum variance unbiased estimators in this case. The purpose of this paper is to exhibit a set of sufficient statistics for the general two-way classification model with unequal numbers in the subclasses. In particular, we show that the row totals, the column totals, and the intra-block error form a set of *sufficient* statistics for the variance components in a two-way classification model with unequal numbers in the subclasses.

35. Minimal Sufficient Statistics for Eisenhart's Model II in a Class of Two-Way Classification Models. David L. Weeks and Franklin A. Graybill, Oklahoma State University. (By title)

The class of designs in which the number of experimental units per block is constant, and the number of observations per treatment is constant, is examined in order to determine a set of minimal sufficient statistics. This class of designs includes as a subset, the balanced incomplete block, and the partially balanced incomplete block designs. Eisenhart's Model II under normal theory is assumed. The number of minimal sufficient statistics is expressed as a function of the distinct characteristic roots of the matrix NN' where N is the incidence matrix of the design. The distribution of each statistic is given and pairwise independence investigated. In the case of the BIB and GD-PBIB's, the statistics are defined explicitly in terms of quantities normally calculated in the analysis of variance. Instructions as to how the statistics may be computed easily for the case of the GD-PBIB's is also given.

36. Two New Continuous Sampling Plans. John S. White, General Motors Technical Center Research Labs.

Two new continuous sampling plans are proposed. Both plans are variations of the concepts used by Dodge and Torrey (*Ind. Qual. Cont.*, Jan., 1951) in CSP-2 and CSP-3. CSP-3 differs from CSP-2 in that following the discovery of a defective unit during sampling inspection, the next four units submitted must be inspected and found non-defective if sampling inspection is to continue. A new plan, called CSP-2.1, is proposed which requires only that the next unit after the defective pass inspection rather than the next four, in order that sampling inspection be continued. In this notation, the original CSP-2 might be denoted as CSP-2.0 and CSP-3 as CSP-2.4. A second plan is given which does require the inspection of four units following the discovery of a defective during sampling inspection but which eliminates the spacing number (i.e. $k = 0$). Graphs giving contours of constant sampling frequency in the $AOQL$ and i = clearance number plane are provided for both plans.

37. On a Class of Covariance Kernels Admitting a Power Series Expansion (Preliminary Report). N. Donald Ylvisaker, Columbia University.

Let $\mathcal{K}(T)$ denote the class of covariance kernels defined on $T \times T$ and let $a = (a_0, a_1, \cdots)$ be a sequence of nonnegative real numbers. The mapping $\varphi_a: K \to \sum_{j=0}^{\infty} a_j K^j$ maps $B_a = \{K \varepsilon \mathcal{K}(T) | \sum_{j=0}^{\infty} a_j K^j(s, s) < \infty \text{ for all } s \varepsilon T\}$ into $\mathcal{K}(T)$. This paper discusses these mappings and in particular the reproducing kernel space associated with the kernel $\varphi_a(K)$ is studied relative to the reproducing kernel space associated with the kernel K. Some applications of these results are noted in reference to problems of mean value estimation under the model $Y(t) = m(t, \beta) + X(t)$, $t \varepsilon T$, $\beta \varepsilon \Lambda \subset R_1$, where $X(\cdot)$ is assumed to be Gaussian process with mean function zero and known covariance kernel.

38. A Calculus for Factorial Arrangements. M. Zelen and B. Kurkjian, National Bureau of Standards and Diamond Ordnance Fuze Laboratories.

A calculus complete with special axioms, operations, and rules of formation is formally defined with respect to factorial arrangements. The object of this calculus is to permit easy manipulation of complicated mathematical operations. Its use enables large order matrix operations to be carried out using logical operations.

CORRECTION TO ABSTRACTS

"Semi-Markov Processes: Countable State Space" and "Stationary Probabilities for a Semi-Markov Process with Finitely Many States

BY RONALD PYKE

Columbia University

The titles of the above-named abstracts, numbers 55 and 74 on pages 240 and 245–46 respectively in the March 1960 *Ann. Math. Stat.*, were reversed in printing. Therefore, "Semi-Markov Processes: Countable State Space" applies to abstract 55 and "Stationary Probabilities for a Semi-Markov Process with Finitely Many States" applies to abstract 74.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Dr. Leo Aioian, recently Head of the Mathematics Section and Senior Mathematical Consultant at Hughes Aircraft Company, has accepted a position as member of the Technical Staff at Space Technology Laboratories.

Oskar N. Anderson, the University of Munich, Germany died on February 12, 1960.

George E. P. Box has left his post as Director of the Statistical Techniques Research Group at Princeton University to take up an appointment at the University of Wisconsin as Professor of Statistics.

Mr. Roshan L. Chaddha has completed his Ph.D. degree at Virginia Polytechnic Institute. He is joining the Department of Statistics at Kansas State University and will be located at Manhattan, Kansas during the next year. The dissertation topic for the Ph.D. degree was "An Inventory Control Problem with Regular and Emergency Demands".

Georges Darmois, professor at the Institut de Statistique of the University of Paris, died on January 3, 1960.

Norman R. Draper has left Imperial Chemical Industries (Plastics Division) England to spend the academic year 1960–61 at the Army Mathematics Research Center, Madison, Wisconsin.

Dr. Seymour Geisser has returned to the National Institute of Mental Health, Washington, D. C. after having served as visiting Associate Professor at the Iowa State University during the Spring Quarter, 1960.

H. S. Graf has joined The Teleregister Corporation, Stanford, Connecticut.

Professor E. J. Gumbel (Columbia University) will participate at the June meeting of the International Statistical Institute in Tokyo and will give papers on the theory of extreme values at the Universities of Kyoto, Osaka, Tokyo (Institute of Technology), and Manila. During July he will give a course on mathematical statistics at the Chulalongkorn University, Bangkok, Thailand.

Dr. John Gurland, Professor in the Statistical Laboratory and Department of Statistics, Iowa State University, has been awarded a travel grant by the Committee on International Conference Travel Grants of the American Statistical Association, to attend the Biometric Society Symposium on Quantitative Methods in Pharmacology at the University of Leyden (Netherlands) May 10–13. Dr. Gurland will act as the official representative of the American Statistical Association at this Symposium, and will also present a paper entitled "Determination of Minute Insecticidal Residues Through Biological Assay."

Stuart T. Hadden has joined the Process Engineering Department of The Dow Chemical Company as a Systems Engineer.

F. M. Hemphill, Ph.D., formerly Professor of Public Health Statistics, School

of Public Health, The University of Michigan, was recently commissioned in the Regular Corps United States Public Health Service. He is now on duty as Scientist Director serving as Chief of the Statistical Design and Analysis Section of the Statistics and Analysis Branch of the Division of Research Grants, National Institutes of Health, Bethesda, Maryland.

Mark L. Hinkle, Jr., is currently a member of the Reliability and Maintainability Unit in General Electric Company's Light Military Electronics Department, 901 Broad Street, Utica, New York.

Palmer O. Johnson, professor of education and chairman of the department of statistics at the University of Minnesota, died at the age of 68 on January 24, 1960.

Shriniwas Keshavarao Katti completed requirements for the Ph.D. Degree in statistics at the Iowa State University in January 1960. He has joined the staff of the new Department of Statistics at the Florida State University, Tallahassee, Florida, as Assistant Professor of Statistics.

Dr. J. H. B. Kemperman, who was engaged last year at the University of Amsterdam (Netherlands), has been promoted to Professor in the Department of Mathematics and Statistics at Purdue University, Indiana.

A. I. Khinchin died at the age of 65 on November 18, 1959. Khinchin had studied and taught at the University of Moscow and V. A. Steklov Mathematical Institute.

Gilbert Lieberman, formerly a mathematician with the Naval Ordnance Laboratory, Silver Spring, Maryland, is now a Senior Engineer with The Radio Corporation of America, Camden, New Jersey.

John W. Mayne, since March 1959, has been Chief of the Operational Research Section at Supreme Headquarters Allied Powers Europe (SHAPE), Paris, France. This section is part of the SHAPE Air Defence Technical Centre's System Evaluation Group, and is concerned mainly with air defense problems of Allied Command Europe.

Robert H. Morris has been named associate director of the newly formed Business Operations Analysis Staff at Eastman Kodak Company, Rochester 4, New York, whose purpose is applying scientific and mathematical techniques as aids in analyzing a wide range of business problems.

Dr. Stanley W. Nash of the University of British Columbia has been appointed as Visiting Associate Professor at the Statistical Laboratory, Iowa State University for a period of one year beginning July 1, 1960.

José R. Padró, Assistant Professor at the University of Puerto Rico, received a Ph.S. in Mathematics from St. Louis University in June, 1960. His dissertation was written under the direction of Dr. Waldo A. Vezeau. Dr. Padró has been on leave-of-absence and sponsored by the University of Puerto Rico while doing his graduate studies. He will resume teaching at the Department of Mathematics, University of Puerto Rico, Río Piedras, Puerto Rico.

B. E. Phillips has been made Assistant Technical Director, Reliability, Ground Support Equipment in The Martin Company, Baltimore, Maryland.

Donald M. Roberts has received his Ph.D. Degree at Stanford University and has accepted an assistant professorship in the Mathematics Department at the University of Illinois.

Ernest M. Scheuer, Space Technology Laboratories, Inc., has received the Ph.D. in mathematics from U. C. L. A.

Earl A. Thomas, formerly Technical Advisor, Ballistic Missiles Division, Burroughs Corporation, has joined the staff of The Institute for Defense Analysis.

Vernon E. Weckwerth has returned as lecturer and Administrative Director of the third Graduate Session of Statistics in the Health Sciences at the University of Minnesota. Mr. Weckwerth was head of Research and Statistics for the American Hospital Association in Chicago and Assistant Director of the Hospital Research and Educational Trust. He also taught at Northwestern University last fall.

NEW MEMBERS

The following persons have been elected to membership in the Institute

Andrews, Horace P., Ph.D. (Pennsylvania State University); Head Statistics Division-Research Laboratories, *Swift and Company, U. S. Yards, Chicago 9, Illinois.*

Balintfy, Joseph L., Diplomas in English and Economics, (University of Technical Sciences, Budapest); Assistant Professor, *Tulane University, School of Business Administration, New Orleans 18, Louisiana.*

Baxter, Colin Benjamin, B.S. (University of Sheffield, England); Lecturer in Mathematics, Harrow Technical College, Watford Road, Northwick Park, Harrow-on-the-Hill, Middlesex, England; *25, Swallowbeck Avenue, Lincoln, England.*

Beale, Evelyn M. L., B.A. (Cambridge University); Member of Mathematics Group, *Admiralty Research Laboratory, Teddington, Middlesex, England.*

Berens, Alan Paul, M.S. (Purdue University); Graduate Assistant, *Purdue University, West Lafayette, Indiana.*

Blischke, Wallace R., M.S. (Cornell University); National Science Foundation Cooperative Graduate Fellow, *Cornell University, Department of Plant Breeding, Ithaca, New York.*

Chakravarti, Indra Mohan, D. Phil. (Sc). (University of Calcutta); Visiting Assistant Professor, *University of North Carolina, Department of Statistics, Chapel Hill, North Carolina.*

Chow, Yuan S., Ph.D. (University of Illinois); Research Staff Mathematician, *I.B.M. Research Center, Yorktown Heights, New York.*

Collins, Gwyn, B.Sc. (Nottingham); Research Associate, Advertising Research Foundation, 3 E. 54th Street, New York City 22, New York; *63 E. 9th Street, New York City 3, New York.*

Dagen, Herbert B., B.A. (City College of New York); Chief, Statistical Operations Division, U. S. Army Chemical Corps, Quality Assurance Technical Agency, Army Chemical Center, Maryland; *3710 Bowers Avenue, Baltimore 7, Maryland.*

Davidson, Harold, Ch. E. (Columbia University); Staff Engineer, I. B. M. Corporation-IPC-ASDD, Yorktown Heights, New York; *455 E. 14th Street, New York 9, New York.*

Farquhar, Thomas H., S.B. (Massachusetts Institute of Technology); Student, Massachusetts Institute of Technology. Cambridge 39, Massachusetts; *22 Magazine Street, Cambridge 39, Massachusetts.*

Foster, William F., B.S. (U. S. Naval Academy); Lieutenant United States Navy, Student, Postgraduate School, Monterey, California, *830 Arlington Place, Monterey, California.*

Gnedenko, Boris, Member Academie of Science of Ukrainien SSR; Professor of Mathematics, Chief Statistical Department, Mathematical Institute of Academie of Science, Kalinine Pl. 6, Kiev, USSR; *Sverdeor Str. 10, App 2g, Kiev 5, USSR.*

Gulotta, Charles William, A.B. (Hunter College); Programmer, Great American Insurance Company, 99 John Street, New York, N. Y., *95-35 114th Street, Richmond Hill 18, New York.*

Hora, Rajinder Bir, M.S. (Panjab University, India); Associate Engineer, Boeing Airplane Company and Graduate Student at University of Washington, Boeing Airplane Company, Renton, Washington; *4738 - 16th North East, Seattle 5, Washington.*

Kappel, Joseph George, M.S. (University of Illinois); Assistant, Department of Mathematics, University of Illinois, Urbana, Illinois; *305 S. Urbana Avenue, Urbana, Illinois.*

Katti, Shriniwas K., Ph.D. (Iowa State University); Assistant Professor, *Florida State University, Tallahassee, Florida.*

Kenworthy, Orville O., M.S. (Oklahoma State); Administrative Assistant, *Ferro Corporation, 4150 E. 56th Street, Cleveland 5, Ohio.*

King, R. Maurice, Jr., B.S. (University of North Carolina); Experimental Statistician, American Cyanamid Company, Stamford Labs, 1937 W. Main Street, Stamford, Connecticut; *52 Sinawoy Road, Cos Cob, Connecticut.*

Kirchgässner, Klaus, Dr. rer. net., (Nat.-Math. Fakultät der Universität Freiburg i.Br., Deutschland); Scientific Assistant, *Institut für Angewandte Mathematik der Universität Freiburg, i.Br., Freiburg i.Br., Deutschland, Friedrichstr. 37.*

Ku, Hsien H., M.S. (Purdue University); Mathematician, National Bureau of Standards, Department of Commerce, Washington, 25, D. C.; *5430-30th Place, N.W., Washington, D. C.*

Laue, Richard V., M.S. (Rutgers University); Statistician, *Bell Telephone Laboratories, Murray Hill, New Jersey.*

Lee, Ray H., M.S. (Stanford University); Chief Mathematician, *Autometric Corporation, 1501 Broadway, New York City, New York.*

Levin, Morris J., Ph.D. (Columbia University); Engineering Scientist, Missile Electronics and Controls Division, Radio Corporation of America, Burlington, Massachusetts; *370 Concord Avenue, Cambridge, Massachusetts.*

McGuire, Charles Bartlett, M.A. (University of Chicago); Economist, *Rand Corporation, 1700 Main Street, Santa Monica, California.*

Mandis, George A., B.S. (Roosevelt University of Chicago); Member Operations Research Analysis, Grumman Aircraft Engineering Corporation, Bethpage, L. I., New York; *George A. Mandis and Associates, 6 Wooleys Lane, Great Neck, New York.*

Marques Henriques, José Manuel, (Lisbon School of Economics); Student, Lisbon School of Economics (Instituto Superior de Ciencias Economicas e Financeiras), R. do Quelhas, 6, Lisbon, Portugal; *R. do Sol, ao Rato, 57, 2° Esq°, Lisbon 2, Portugal.*

Mehr, Cyrus B., M.S. Industrial Engineering (Purdue University); Instructor and part-time Student, Purdue University, West Lafayette, Indiana; *207-11 Airport Road, West Lafayette, Indiana.*

Mohanty, Sri Gopal, M.A. (Punjab University, India); Graduate Student, *Department of Mathematics, University of Alberta, Edmonton, Canada.*

Montzingo, Lloyd J., Jr., M.A. (University of Buffalo); Instructor, *University of Buffalo, Buffalo 14, New York.*

Murray, Charles W. Jr., B.A. (Duke University); Engineer, Melpar, Inc., 3000 Arlington Blvd., Falls Church, Virginia; *109 Chapel Drive, Annandale, Virginia.*

Neuts, Marcel Fernand, M.S. (Stanford University); Graduate Student, Department of Statistics, Stanford University, Stanford, California; *2365 Waverly Street, Palo Alto, California.*

Nielsen, Aage Volund, M.S. Chem. Eng. (Technical University of Denmark); Civilingenior,

Statistical Institute of University of Copenhagen, Denmark; *Norre Alle 75, Room 544, Copenhagen O, Denmark.*

Novick, Melvin R., M.A. (Roosevelt University); Student Research Assistant, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina; *102C Isley Street, Chapel Hill, North Carolina.*

Okana, K. Frederick, M.A. (University of Minnesota); Mathematical Statistician, *National Aeronautics and Space Administration (AAR) 1520 H. Street, N.W., Washington 25, D. C.*

Pandharipande, Vikas Raghunath, Int. in Science (Nagpur University, India); *℅ Mrs. K. Pandharipande, Advocate, Ramdas Peth, Nagpur, India.*

Posten, Harry O., M.S. (Kansas State University); Ph.D. Candidate, *Virginia Polytechnic Institute, Blacksburg, Virginia.*

Rao, U. V. Ramamohana (U. V. R.) M.A. (Andhra University, India); Graduate Assistant, *Department of Mathematics, Indiana University, Bloomington, Indiana.*

Ray, Sudhindra Narayan, M.S. (Calcutta University, India); Student, University of North Carolina, Chapel Hill, N. C., Department of Statistics; *308 Connor Dorm., University of North Carolina, Chapel Hill, N. C.*

Richardson, Earle Wesley, Jr., A.B. (Georgetown University); Student, American University, Washington 16, D. C.; *3109 44th Street, N.W., Washington 16, D. C.*

Schwarz, Gideon E., M.Sc. (Hebrew University); Graduate Student, *Department of Mathematical Statistics, Columbia University, New York 27, N. Y.*

Shy, William H., M.A. (University of Georgia); Research Analyst, Kimberly-Clark Corporation, Neenah, Wisconsin; *P. O. Box 22, Neenah, Wisconsin.*

Singh, Shorh Nath, M.A. (Banaras Hindu University, India); Graduate Student, *Department of Statistics, University of California, Berkeley, California.*

Soriano, Abraham, B.S. (Rensselaer Polytechnic Institute); Industrial Engineering Department, Johns Hopkins University (Hospital), Baltimore, Md.; *3301 Saint Paul Street, Baltimore 18, Maryland.*

Sternberg, I. Paul, M.S. (Rutgers University); *Director, Quality Control, Whittaker Gyro, Division of Telecomputing Corporation, 16217 Lindbergh Street, Van Nuys, California.*

Sukhatme, (Mrs.) Shashikala, M.S. (University of Poona, India); Graduate Assistant, *Department of Statistics, Michigan State University, East Lansing, Michigan.*

Susco, Dante V., M.A. (University of California, Los Angeles); Staff Member, *Los Alamos Scientific Laboratory, P.O. Box 1663, Los Alamos, New Mexico.*

Umegaki, Hisaharu, Ph.D. (Mathematical Institute, Tohoku University, Japan); Assistant Professor, Department of Mathematics, Tokyo Institute of Technology, Oh-okayama, Meguro-ku, Tokyo, Japan; *Akasaka Aoyama Minamicho 6-108, Minato-ku, Tokyo, Japan.*

Van Dyke, John, M.A. (Michigan State University); Assistant Instructor, *Department of Statistics, Michigan State University, East Lansing, Michigan.*

Wolfe, John H., B. S. (California Institute of Technology); Student, Department of Psychology, University of California, Berkeley, California; *2299 Piedmont, Berkeley 4, California.*

Wu, Shih Yen, M.A. (Northwestern University); Assistant Professor of Economics, *Los Angeles State College, Los Angeles 32, California.*

PRELIMINARY ACTUARIAL EXAMINATIONS PRIZE AWARDS ANNOUNCED

The winners of the prize awards offered by the Society of Actuaries to the nine undergraduates ranking highest on the score of the General Mathematics Examination of the 1960 Preliminary Actuarial Examinations are as follows:

First Prize of $200

Gitlin, Todd A. Harvard University

Additional Prizes of $100 each

Emerson, William R.	California Institute of Technology
Goodman, Richard H.	Harvard University
Landman, Maurice A.	Harvard University
Lorden, Gary A.	California Institute of Technology
McDonnell, Robert N.	University of Chicago
Newmeyer, John A.	California Institute of Technology
Sampson, Schuyler S.	Bowdoin College
Shulsky, Abram N.	Cornell University

The Society of Actuaries has authorized a similar set of nine prizes for 1961. Beginning in 1961, the Preliminary Actuarial Examinations will consist of two examinations: The General Mathematics Examination (based on the first two years of college mathematics), and The Probability and Statistics Examination. The 1961 Preliminary Actuarial Examinations will be prepared by the Educational Testing Service under the direction of a committee of actuaries and mathematicians, and will be administered by the Society of Actuaries at centers throughout the United States and Canada on November 16, 1960 and on May 10, 1961. The closing date for applications is April 1, 1961. Further information concerning these Examinations can be obtained from the Society of Actuaries, 208 South LaSalle Street, Chicago 4, Illinois.

RESEARCH FELLOWSHIPS IN PSYCHOMETRICS OFFERED

The Educational Testing Service is offering for 1961–62 its fourteenth series of research fellowships in psychometrics leading to the Ph.D. degree at Princeton University. Open to men who are acceptable to the Graduate School of the University, the two fellowships each carry a stipend of $3,750 a year, plus an allowance for dependent children. These fellowships are normally renewable. Fellows will be engaged in part-time research in the general area of psychological measurement at the offices of the Educational Testing Service and will, in addition, carry a normal program of studies in the Graduate School.

Suitable undergraduate preparation may consist either of a major in psychology with supporting work in mathematics, or a major in mathematics together with some work in psychology. However, in choosing fellows, primary emphasis is given to superior scholastic attainment and research interests rather than to specific course preparation.

The closing date for completing applications is January 6, 1961. Information and application blanks will be available about September 15 and may be obtained from: Director of Psychometric Fellowship Program, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey.

INSTITUTE OF MANAGEMENT SCIENCES HOLDS MEETING

The Seventh International Meeting of the Institute of Management Sciences will be held at Hotel Roosevelt, New York City, October 20–22, 1960. Among

the topics for sessions of the meeting are the following: information processing and management science, computors and simulation techniques, and mathematical methods for management science. Further information may be obtained from Mr. James Townsend, Union Carbide Corp., 270 Park Avenue, New York 17, N. Y.

SYMPOSIUM ON MATHEMATICAL OPTIMIZATION TECHNIQUES

A Symposium on Mathematical Optimization Techniques will be held on October 18, 19, and 20, 1960 at the University of California, Berkeley. The Symposium is sponsored jointly by the University of California and RAND Corporation, with twenty invited speakers presenting papers during the three-day sessions on such topics as linear, non-linear and discrete programming, variational processes: adaptive and stochastic, optimal decision processes, and optimum networks and structures. For further information, write to Robert M. Oliver, Department of Industrial Engineering, University of California, Berkeley 4, California.

UNIVERSITY OF MINNESOTA STATISTICS DEPARTMENT ESTABLISHED

During the academic year 1958–1959 the Department of Statistics was established at the University of Minnesota. The Department has supplemented and coordinated the statistical activities of the University—graduate curriculum, research, and consulting. The Department's organization involves direct appointments as well as joint appointments in mathematics and the sciences. Following is the current staff of the Faculties of Statistics: *Statistics*: L. Hurwicz, I. Olkin, D. Richter, I. R. Savage, M. Sobel; *Mathematics*: G. Baxter, M. Donsker, B. Lindgren, S. Orey, W. Pruitt, E. Reich, F. Spitzer; *Agriculture*: R. Comstock, C. Gates; *Biostatistics*: J. Bearman, J. Berkson, B. Brown, E. Johnson, R. McHugh; *Business Administration*: D. Hastings, J. Neter; *Industrial Engineering*: G. McElrath.

DOCTORAL DISSERTATIONS IN STATISTICS, 1959

The following listing was inadvertently omitted from the June 1960 issue of the *Annals*:

William Leonard Harkness, Michigan State University, major in statistics, "An Investigation of the Power Function for the Test of Independence in 2×2 Contingency Tables."

REPORT OF THE NEW YORK CITY MEETING OF THE INSTITUTE OF MATHEMATICAL STATISTICS

The eighty-fourth meeting of the Institute of Mathematical Statistics was held at Teachers College, Columbia University, on April 21–23, 1960, in conjunction with a meeting of the Biometric Society (Eastern North American Region). Program chairman for the meeting was Rosedith Sitgreaves, Teachers College; Ronald Pyke, Columbia University, was Assistant Secretary. In addition to technical sessions there were an informal party at the Columbia University Men's Faculty Club, and two coffee hours.

211 persons, including 142 members of the Institute registered for the meeting. The program of the meeting was as follows. All sessions for invited papers were jointly sponsored by the Institute and the Biometric Society.

THURSDAY, APRIL 21, 1960

10:00–12:00 a.m.—Inventory and Queueing Theory

Chairman: Howard Levene, Columbia University.

1. *"A Multi-Server Queueing Problem,"* Peter E. Ney, Cornell University.
2. *"Queues in Series,"* Jerome Sacks, Columbia University.
3. *"On the Transient Behavior of a Queueing Process with Batch Service,"* Lajos Takacs, Columbia University.

1:30–2:30 p.m.—Invited Address

Chairman: Boyd Harshbarger, Virginia Polytechnic Institute.

1. *"Two Methods of Constructing Exact Tests,"* James Durbin, University of North Carolina.

2:30–4:30 p.m.—Selected Topics I

Chairman: C. Y. Kramer, Virginia Polytechnic Institute.

1. *"On Comparing Different Tests of the Same Hypothesis,"* H. A. David and Carmen A. Perez, Virginia Polytechnic Institute.
2. *"On the Replacement of Periodically Inspected Equipment,"* C. Derman, Columbia University.
3. *"Lower Bounds on the Probability Associated with Certain Confidence Regions for the Multivariate Median,"* Ernest M. Scheuer, U.C.L.A. and Space Technology Laboratories, Inc.

2:30–4:30 p.m.—Contributed Papers I

Chairman: Milton Sobel, Bell Telephone Laboratories.

1. *"A Noiseless Comma-Free Coding Theorem,"* Thomas S. Ferguson, U.C.L.A. and Princeton University.
2. *"On Centering Infinitely Divisible Processes,"* Ronald Pyke, Columbia University.
3. *"Inference About Non-Stationary Markov Chains,"* Ruth Z. Gold, Columbia University.
4. *"On Linear Estimation of a Single Parameter of a Mean Function under Second Order Disturbance"* (Preliminary Report), N. Donald Ylvisaker, Columbia University.
5. *"Asymptotic Shapes of Optimal Stopping Regions for Sequential Testing,* (Preliminary Report), Gideon Schwarz, Columbia University (Introduced by T. W. Anderson).

6. "*The Asymptotic Power of the Kolmogorov Tests of Goodness of Fit*," Dana Quade, University of North Carolina.

FRIDAY, April 22, 1960

9:00–10:30 a.m.—Problems in Multivariate Analysis

Chairman: I. Blumen, Cornell University.

1. "*Multivariate Analysis in Psychology and Education*," Rolf Bargmann, Virginia Polytechnic Institute.
2. "*Multivariate Experimental Designs*," Harry Rosenblatt, Federal Aviation Agency.

9:00–10:30 a.m.—Contributed Papers II

Chairman: Marvin Zelen, National Bureau of Standards.

1. "*Efficient Sequential Estimators with High Precision Only in a Small Interval*," Allan Birnbaum, New York University.
2. "*Asymptotic Variance as an Approximation to Expected Loss for Maximum Likelihood Estimates*," William D. Commins, Jr., Alexandria, Va.
3. "*Extensions of the Poisson and the Negative Binomial Distribution*," A. Clifford Cohen, Jr., University of Georgia.
4. "*Concerning Achievement of the Lower Bound for the Power of the Kolmogorov-Smirnov Test of Fit*," Judah Rosenblatt, Purdue University.

11:00–1:00 p.m.—Some Problems in the Analysis of Variance

Chairman: Donald A. Gardiner, Oak Ridge National Laboratory.

1. "*Analysis of Variance of Variances*," Howard Levene, Columbia University.

Discussant: Robert E. Bechhofer, Cornell University.

2. "*Properties of Model II and of Mixed Models*," Leon Herbach, New York University.

Discussant: Jack Nadler, Bell Telephone Laboratories, Whippany, N. J.

2:30–4:30 p.m.—Problems in Medical Statistics

Chairman: John W. Fertig, School of Public Health, Columbia University.

1. "*A New Closed Sequential Scheme for Clinical Trials*," Clive C. Spicer, Imperial Cancer Research Fund and National Institutes of Health.
2. "*Some More Difficulties in Clinical Trials*," Donald Mainland, New York University, College of Medicine.
3. "*Effects of Non-Linearity on Subjective and Objective Analyses of Hormonal Assay Data*," Nathan Mantel, National Institutes of Health.

2:30–4:30 p.m.—Contributed Papers III

Chairman: J. A. Zoellner, General Electric Co.

1. "*A Robust Approximate Confidence Interval for Components of Variance*," Howard Levene, Columbia University.
2. "*The Partition of Phenotypic Variance Based on the Genic-Environmental Interaction Model*," Cecil L. Kaller and Virgil L. Anderson, Purdue University.
3. "*Power Functions for the Test of Independence in 2×2 Contingency Tables*," William Harkness, Pennsylvania State University.
4. "*On Dependent Tests in Analysis of Variance*," S. N. Roy and P. R. Krishnaiah, University of North Carolina.
5. "*On the Admissibility of a Class of Tests in Normal Multivariate Analysis*," S. N. Roy and W. F. Mikhail, University of North Carolina.

6. *"On a Generalization of Balanced Incomplete Block Designs,"* J. N. SRIVASTAVA AND S. N. ROY, University of North Carolina.
7. *"Partially Balanced Arrays,"* I. M. CHAKRAVARTI, University of North Carolina and Indian Statistical Institute (Introduced by David B. Duncan).
8. *"Multi-Stage Bayesian Lot-by-Lot Sampling Inspection,"* HERBERT B. EISENBERG, System Development Corporation (Introduced by Herbert T. David).

SATURDAY, APRIL 23, 1960

9:00–10:00 a.m.—Contributed Papers IV

Chairman: D. B. DUNCAN, University of North Carolina.

1. *"Maximum Likelihood Characterization of the Normal Distribution,"* HENRY TEICHER, Purdue University.
2. *"Invariant Bayes Rules* (Preliminary Report), MORRIS SKIBINSKY AND KENZO SEO, Purdue University.
3. *"A Central Limit Theorem for Systems of Regressions,"* EDWARD J. HANNAN, University of North Carolina (Introduced by David B. Duncan).
4. *"Some Results on Error Correcting Non-Binary Codes,"* D. K. RAY-CHAUDHURI, Case Institute of Technology.

10:00–11:00 a.m.—Invited Address

Chairman: SEBASTIAN LITTAUER, Columbia University.

1. *"Sampling Inspection, Prior Distributions and Cost,"* A. HALD, University of Copenhagen and Princeton University.

11:00–1:00 p.m.—Selected Topics II

Chairman: J. WOLFOWITZ, Cornell University.

1. *"Estimation of the Correlation of Non-Stationary Random Functions,"* J. KAMPÉ DE FERIET, University of Lille and Harvard University.
2. *"Statistical Sufficiency and the Foundations of Thermodynamics,"* B. MANDELBROT, IBM Research Center.
3. *"Markov Renewal Processes of Zero Order,"* RONALD PYKE, Columbia University.

Contributed Papers Presented by Title

1. *"Transition Probabilities for Telephone Traffic,"* V. E. BENEŠ, Bell Telephone Laboratories and Dartmouth College.
2. *"A Representation of the Bivariate Cauchy Distribution,"* THOMAS S. FERGUSON, U.C.L.A. and Princeton University.
3. *"On Two Methods of Unbiased Ratio and Regression Estimation,"* W. H. WILLIAMS, McMaster University, Ontario.
4. *"A Generalization of a Theorem of Balakrishnan,"* N. DONALD YLVISAKER, Columbia University.
5. *"An Inequality for Balanced Incomplete Block Designs,"* W. F. MIKHAIL, University of North Carolina.

PUBLICATIONS RECEIVED

Annuario Estadistico de España, Edicion Manual, Presidencia del Gobierno, Instituto Nacional de Estadistica, Ferraz 41, Madrid, Spain, 1960, 1001 pp.

Cowden and Cowden, *Practical Problems in Business Statistics*, 2nd ed., Prentice-Hall, Inc., New York, 1960, 95 pp.

Goldberg, Samuel, *Probability: An Introduction*, Prentice-Hall, Inc., New York, 1960, 238 pp., $5.95.

Kellerer, Hans, *Statistik im Modernen Wirtschafts-und Sozialleben* (in German), Rowahlt, Hamburg, 1960, paperback, 289 pp.

Lukacs, E., *Characteristic Functions*, No. 5 of Griffin's Statistical Monographs and Courses, ed. M. G. Kendall, Hafner Pub. Co., New York, 1960, 216 pp., $6.45.

Ingram Olkin, et al., *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, Stanford University Press, Stanford, 1960, 517 pp., $6.50.

Quenouille, M. H., *Rapid Statistical Calculations: a Collection of Distribution-Free and Easy Methods of Estimation and Testing*, Hafner Pub. Co., New York, 1959, 35 + pp., $1.65.

Robinson, Enders A., *An Introduction to Infinitely Many Variates*, No. 6 of Griffin's Statistical Monographs and Courses, ed. M. G. Kendall, Hafner Pub. Co., New York, 1960, 132 pp., $4.75.

United Nations Statistical Yearbook, United Nations, New York, 1960, 618 pp., cloth $10.00, paper $8.00.

Weissberg, Alfred and Beatty, Glenn H., *Tables of Tolerance-Limit Factors for Normal Distributions*, Battelle Memorial Institute, Columbus, Ohio, 1960, 28 pp. Individual copies are available without charge, while the supply lasts, from the Battelle Publications Office, 505 King Avenue, Columbus 1, Ohio.

THE INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

OFFICERS

President:

E. L. Lehmann, Department of Statistics, University of California, Berkeley 4, California

President-Elect:

A. H. Bowker, Department of Statistics, Stanford University, Stanford, California

Secretary:

G. E. Nicholson, Jr., Department of Statistics, University of North Carolina, Chapel Hill, North Carolina

Treasurer:

Gerald Lieberman, Department of Statistics, Standard University, Stanford, California

Editor:

William Kruskal, Department of Statistics, Eckhart Hall, University of Chicago, Chicago 37, Illinois

The purpose of the Institute of Mathematical Statistics is to encourage the development, dissemination, and application of mathematical statistics.

Membership dues including a subscription to the ANNALS OF MATHEMATICAL STATISTICS are $10.00 per year for residents of the United States or Canada and $5.00 per year for residents of other countries. There are special rates for students and for some other classes of members. Inquiries regarding membership in the Institute should be sent to the Secretary of the Institute.

Contents (continued)

MEETINGS OF THE INSTITUTE

TENTATIVE SCHEDULE

EASTERN REGIONAL MEETING—
Ithaca, New York, April 21–22, 1961

Abstracts should be submitted in duplicate to the Editor, preferably on abstract blanks, which can be obtained from the IMS Secretary. Abstracts must be received at least 50 days before the first day of the meeting at which they are to be presented, indicating whether presented by title or in person. (Only one contributed paper may be given in person at any one meeting.) They may be printed prior to the publication of the report of the meeting. Those received by April 30 will appear in the September *Annals*, by July 31 in December, etc. Abstracts should be limited to 200 words or the equivalent, and should avoid displayed expressions and complicated formulae. They can be accepted from non-members of the IMS only if transmitted by members.